Supplementary Appendix:

A high level of inter-genera gene exchange shapes the evolution of haloarchaea in an isolated Antarctic lake

Matthew Z. DeMaere, Timothy J. Williams, Michelle A. Allen, Mark V. Brown, John A.E. Gibson, John Rich, Federico M. Lauro, Michael Dyall-Smith, Karen W. Davenport, Tanja Woyke, Nikos Kyrpides, Susannah G. Tringe and Ricardo Cavicchioli

Results and Discussion

Fragment recruitment (FR), highly degenerate regions and reference mapping. 454 sequencing was performed on the 0.1 µm fraction from the 5, 13 and 24 m depths, the 0.8 and 3.0 µm fractions from 24 m, and all 3 fractions pooled together for the 36 m depth (Table S1). Illumina sequencing was performed on the 0.8 and 3.0 µm fractions from the 24 m depth. FR was performed individually on each dataset and carried out against all publically available complete and draft bacterial, archaeal and viral genomes from NCBI, all organelle genomes from EHL and the four DL haloarchaea (tADL, DL31, *Hl* and DL1). The only eucaryal reference examined was a halophilic green alga *Dunaliella salina* CCAP 19/18 (http://genomesonline.org/cgi-bin/GOLD/bin/GOLDCards.cgi?goldstamp=Gi13840).

Genomes were analysed independently using FR-Hit and the resulting combined set of read assignments across all genomes was reduced to a non-redundant set by selection of the single longest and highest identity alignment per read, where alignments were filtered by application of a stringency test (read coverage > 90%, identity > 98%). From eight million 454 Titanium reads, the four DL haloarchaea were the four most abundant organisms recruiting 28.3% of all reads and represent 94.4% of all recruited reads. Raw draft *Dunaliella* sp. sequence recruited 131949 reads, however upon inspection it was clear that recruited reads were almost exclusively aligned to low complexity repeat regions within the draft sequence. After masking low complexity sequence (RepeatMasker v3.3.0 using –noint) recruitment to *Dunaliella salina* dropped to 1857 reads, making a distant 5th in abundance (0.08%). In a further wider step, no reads were recruited to 3900 publically available organelle genomes from EMBL. Recruitment to DL1 primary replicon Contig38 was observed to be concentrated largely (63% of assigned reads) along an 80 kb region (380..460 kb) with mean read-depth of 54.2, outside of this region mean read-depth fell to 2.5 suggesting that the majority of recruiting reads were of other genomic origin. In view of the relative abundance of SSU rRNA gene pyrotag matches to *Dunaliella* 18S and chloroplast 16S rRNA gene sequences, the lack of FR from the metagenomic datasets suggests that DL *Dunaliella* have genomes that are quite different to those chromosome or chloroplast genomes of *Dunaliella* sp. that have been sequenced.

The proportion of mapped reads is very stable for the four dominant organisms, all of which are known to exist within the lake. The tail of weakly recruiting species however, is sensitive to the stringency filter. There is a six-fold increase in the number of weakly recruiting organisms (recruiting < 1%) when lowering the stringency (read coverage > 75%, identity > 80%), while the number of reads recruited increases to 38.2%. In relative proportion, the tail is dominated by haloarchaea assignments (10 of the next 15 organisms) accounting for 82.9% of recruited reads in the tail. The tail included hits to organisms and viruses that are not likely to be present in DL (*e.g.* insect virus). These read assignments provide little value in determining their biological origin, and hence possible contribution to the ecosystem.

For the 0.1 µm filter size the abundance profile has little variation across depths (5, 13 and 24 m) (Fig. S5). For a constant depth (24 m), the 0.8 µm and 3.0 µm fractions compared to 0.1µm were enriched for DL31, indicating DL31 cells in DL are likely to be relatively large thereby partitioning on the larger size filters. The 36 m sample (pooled

size fractions) had a similar profile to the other depths, although the proportion of tADL was somewhat lower. Overall recruitment changed little by depth or filter size across the lake. As DL is effectively homogeneous in community profile reads were pooled for most read-based analyses. This profile for the lake was also reflected in the SSU pyrotag data (Fig. S2-S5).

FR revealed short regions (1-2 kb) of high degeneracy within the metagenome, where local read-depth greatly exceeded the replicon global median. Genomic coordinates of all such degenerate regions were determined by application of peak detection to replicon read-depth traces (read depth > 3-fold median read-depth) and subsequently reduced to a non-redundant set by clustering with CD-hit at 98% identity. Of the initial 197 individual sequences with mean read-depth of 2850, 15 unique sequence clusters were determined with a total extent of 21,747 bp, representing 4.8% (111,095) of all recruited reads. The distribution of occupancy follows an 80/20 power-law with the largest 3-4 clusters comprising 75-80% of extracted sequences. Annotation by BLASTX against Refseq_protein identified 14 of 15 clusters as insertion sequences (ISs), with the remaining single-sequence cluster identified as a conserved hypothetical found in haloarchaea. DL metagenomic reads were subsequently filtered for these highly degenerate sequences and FR repeated. The resultant recruitment plots (Fig. 1 and Fig. S7) show pronounced holes in recruitment depth, which also coincide with those from gsMapper derived reference mapping (detailed below).

A bipartite association network of degenerate cluster to replicon helps to show that the four largest clusters cl_8 (ISHla1), cl_1 (ISHla6), cl_0 (ISHla7) and cl_6 (ISDL31_7; newly defined in DL) are highly connected across the replicons of all four isolate genomes (out degree: 8, 5, 7, 4 respectively) (Fig. 3). Though no cluster is fully associated with all nine replicons, taken at the genome level, clusters cl_8 and cl_0 are fully associated with all four genomes. Considering the network in reverse perspective, tADL, *Hl* and DL31 are the most highly connected genomes (weighted in-degree: 58, 44, and 43) while DL1 (weighted in-degree: 14) primarily associates with cl_8 and only weakly with cl_0 and cl_6. Note that for clarity the Fig. S13 has been filtered to exclude weak (low weight) edges (e.g. between DL1 and cl 0 and cl 6).

Ranked by decreasing weighted in-degree, genome order is conserved as cluster nodes are deleted from the network in order of increasing out-degree, changing dramatically only when reduced to a single node (cl_8). This highlights the relative strength of association between tADL and the second largest cluster cl_1; three-fold greater than the other three genomes combined associative weight with cl_1.

By weighted in-degree, the tADL main chromosome (WID: 61, Contig32) possesses 31% of all edges in the network, making it both the most significant primary replicon and most significant replicon overall, followed by secondary replicons from DL31 (WID: 20, Contig115) and *Hl* (WID: 17, NC_012030). Secondary replicons from DL31 and *Hl* possess approximately two-fold more edges than their respective main chromosomes. Despite having no such secondary replicon, the most abundant organism tADL manages to accommodate the community majority of ISs.

As 70% of reads were unassigned in FR, a question addressed was whether the very high read-depth of degenerate regions was accounted for simply by the combined contribution of the four isolate genomes. A self-consistency check of observed cluster read-depth was obtained by the linear combination,

$$d_{pred} = \sum_i n_i d_i$$
$$\Delta d = d_{pred} - d_{obs}$$

where n_i is per-replicon cluster copy number, d_i the robust estimate of replicon read-depth, and i iterates over all replicons.

Although there is some discrepancy between observation and prediction for cl_8 and cl_1, the majority of observed read-depth is explained by copy number and replicon superposition. For cl_0 however, roughly 40% of observed read-depth was not explained by replicon superposition and was attributable to a 5th genome (see **Whole metagenome assembly** below).

Reference mapping was performed with GS Reference Mapper v2.6 for each genome against all DL samples. Estimated distributions of read-depth for each replicon show that primary replicons for DL31, *Hl* and tADL (Contig115, NC_012029 and Contig32) are smooth and unimodal with median values of 74-, 36- and 221-fold, while as seen in FR, DL1 only maps reads in significant depth in one 80 kb region. Secondary replicon read-depth distributions appear as a superposition of multiple stochastic processes, where regions of degenerate sequence cause high levels of variability (Fig. 1 and Fig. S7).

SNP and recombination analysis. Ninety million reads comprised of two paired-end Illumina lanes from 24 m depth (0.8 um and 3.0 um) were aligned with Bowtie to the four DL haloarchaea genomes following recent methodology (1). Of the 45.6 million reads (49.5%) which were mapped; 50.9% were assigned to tADL, 29.8% to DL31, 16.3% to *Hl* and 3.0% to DL1. Mean recruited read-depths for primary replicons were tADL 530, DL31 301, *Hl* 112, and DL1 14. In the case of DL1, mean read-depth is a deceiving measure as nearly all reads were recruited to a single short region (380 kb..460 kb) while elsewhere read-depth was close to zero.

SNPs with mapped read-depth above 20-fold and variant frequency above 0.9 were considered fixed mutations within the DL population (Table S6). The proportion of fixed SNPs within intergenic or coding regions was in proportion to the number of coding bases in each case (Fisher exact p-values: tADL 0.153, DL31 0.624, *Hl* 0.732, DL1 n/a). Four way contingency tables of reference to variant base demonstrated a bias towards transitions (A ↔ G, C ↔ T) for all main chromosomes (tADL 71.7%, DL31 81.9%, *Hl* 59.2%, DL1 n/a).

High numbers of signal transduction genes (T) and transcriptional regulation genes (K) were reported in *Leptospirillum* populations in acid mine drainage biofilms (1). For the dominant DL organism tADL, the most highly assigned category is “general function prediction only” (R), followed by “energy production and conversion” (C) (Table S7).

Estimating substitution rate,

$$\theta \cong N_{SNP} / N_{gen} / \ell_{genome}$$

Using *Hl* growth rates and DL annual temperature profile (2,3) a power-law regression can be used to model generation time T_{gen} as a function of lake temperature t_{lake}.

$$T_{gen} \cong A t_{lake}^{-2.503}$$

Where A = 43829 and R^2 = 0.99565.

This model predicts that there are only three months in a year (January, February and December) during which time the lake has discernible haloarchaeal growth. Generation times for these productive months (179, 773 and 999 hours respectively) equates to 5.77 generations per year. This is 100-fold fewer generations annually than in the case of acid mine drainage biofilms[1] where T_{gen} = 0.0017 yr $^{-1}$ or 588 generations per year, where the newly emerged bacterial hybrids have demonstrated ecological success.

Because of this slow growth rate in DL, the inferred substitution rate θ of DL haloarchaeal species by this method would be extremely high. Additionally, regions of highly degenerate sequence do not account for the observed fixed number of SNPs.

PIIM v2.02 was used to infer scaled rates of mutation and recombination ($\theta = 2N_e\mu, \rho = 2N_e c$) for each DL isolate replicon. PIIM's pre-generated likelihood lookup table limits read-depth to 50-fold, therefore requiring downsampling of our data. Downsampling by random selection was performed using Picard v1.77 and BAM to ACE format conversion accomplished using Consed v23. Rates were estimated in sliding 50 kb windows and resampling repeated ten times with randomly selected seeds. Confidence intervals (95% CI) for θ, ρ were estimated by bootstrapping in R (N=10,000) using the boot package. Rates estimation in PIIM was performed using the likelihood lookup table for θ = 0.1, since as many as 25% of predictions resulted in $\rho = \infty$ when using the likelihood table for θ = 0.01. DL1 was excluded from this analysis due insufficient coverage.

Confidence intervals for primary replicons were: tADL Contig32 (θ = 0.0088, 0.0091) (ρ = 0.0307, 0.0340); DL31 Contig115 (θ = 0.0258, 0.0264) (ρ = 0.0209, 0.0232); *Hl* NC_012029 (θ = 0.0376, 0.0390) (ρ = 0.0386, 0.0426). Confidence intervals for secondary replicons were: DL31 Contig114 (θ = 0.0224, 0.0290) (ρ = 0.0214, 0.0286); *Hl* NC_012028 (θ = 0.0219, 0.0257) (ρ = 0.0242, 0.341); *Hl* NC_012030 (θ = 0.0238, 0.0267) (ρ = 0.0192, 0.0282). As the length of DL31 Contig113 was less than 50kb, only 10 estimations were determined for each parameter and consequently only means were determined (θ = 0.0208 ± 0.0008) (ρ = 0.0434 ± 0.0043).

The predicted range of the ratio of rates (ρ/θ) suggests that genetic variability within tADL (3.37,3.86) occurs at a ~4-fold higher rate by recombination than mutation, whereas for the primary replicons in DL31 (0.79,0.90) and *Hl* (0.99,1.13) the suggestion is that the rates are approximately equivalent sources of variability. For longer secondary replicons (Contig114, NC_012028, NC_012030) the ratio of rates is also close to unity (DL31 Contig114: 0.74, 1.28; DL31 Contig113: 2.09; *Hl* NC_012028: 0.94, 1.56; *Hl* NC_012030: 0.72, 1.18). While the genome variation occurring via ISs and long high identity regions (HIR; see **High identity regions** below) provides information about the extent of genome variation in the DL haloarchaea, these ratio values indicate about the relative contribution of recombination and mutation. The low rate of mutation *vs* recombination in tADL may indicate that the fidelity of point mutation repair is superior in tADL compared to DL31 and *Hl*.

CAI/CBI. Codon adaptation and codon bias indexes (CAI/CBI) for each genome were determined by CodonW, where putative optimal codons sets are determined by a two way Chi-squared contingency test of the two extremes of the principle trend of a correspondence analysis.

As validation, CAI/CBI values of 77 suspected highly expressed proteins in *Hl* (4) were compared against those of the whole main chromosome. Median CAI/CBI for the whole chromosome were 0.620/0.652, while for the 77 highly expressed proteins 0.723/0.781; whole *vs* highly expressed distributions differed significantly (Mann-Whitney two tailed, n_1=2723, n_2=77, CAI U=29352, p-value 2.2e-16 and CBI U=19465 p-value=2.2e-16). Though a clear bias towards optimality is evident, this set of highly expressed proteins could not be used for codon weighting parameter estimation (w_i) due to amino acid under-sampling.

Comparison of putative optimal codon sets between the four genomes shows a high degree of overlap, where 21 of 25 predicted codons (conserved ratio 0.84) are identical. Compared against the archaeal hyperthermophile *Thermoproteus neutrophilus V24Sta*, the number of conserved codons decreases to 18 of 28 (conserved ratio 0.64). Classical multidimensional scaling (MDS) of an Euclidean distance matrix of estimated parameters (w_i) shows DL1 and *Hl* to be the most similar, while DL31 and tADL are increasingly separated and the expected outlier *T. netriphilus* the furthest removed.

Secondary replicons within the DL genomes show consistently lower CAI and CBI than the main replicon. Additionally, the primary replicons of tADL and DL1 possess extended regions of low CAI/CBI scoring genes (Fig. 1 and Fig. S7) some of which also contain HIR (see **High identity regions** below). These regions often coincide with strong variations in FR and increased density of mobile elements. This could be considered as evidence of recent incorporation and high volatility.

Ortholog groups. Thirteen haloarchaeal genomes (*Haladaptatus paucihalophilus* DX253; *Halalkalicoccus jeotgali* B3, DSM 18796; *Haloarcula marismortui* ATCC 43049; *Halobacterium salinarum* R1, DSM 671; *Halobacterium* sp. NRC-1; *Haloferax volcanii* DS2, ATCC 29605; *Halogeometricum borinquense* PR3, DSM 11551; *Halomicrobium mukohataei* arg-2, DSM 12286; *Haloquadratum walsbyi* HBSQ001, DSM 16790; *Halorhabdus utahensis* AX-2, DSM 12940; *Haloterrigena turkmenica* DSM 5511; *Natrialba magadii* ATCC 43099; *Natronomonas pharaonis* Gabara, DSM 2160) were included together with the four DL genomes in an ortholog cluster analysis. Ortholog clusters were determined using OrthoMCL (v2.0.2) where granularity (-I 1.4) was adjusted for best agreement between the resultant clusters and known single copy genes (*e.g.* ribosomal proteins, tRNA synthetases). From 52973 genes there were 6343 predicted ortholog clusters, 723 containing only one gene from each of the 17 genomes, which increased to 889 when constraints were reduced to allow an absence in one genome. Freeing the constraint on copy number to be greater than one, there were 894 clusters conserved in one or more copies per genome across all 17 haloarchaea, which increased to 1062 when constraints were again relaxed to permit one absent genome. Cluster function was inferred by homology to Archaeal COGs, where HMMER version 2.3.2 was used to create profile HMMs for each group from Clustalw2 multiple sequence alignments and ArCOG assignment decided by a simple voting heuristic.

The 894 ortholog clusters shared across all 17 genomes were used to define the core haloarchaeal gene content. The breakdown of functional classification for essential gene content was in agreement with other recent work (5) where 11% are involved in cellular processes and signalling, 30% in information storage and processing, 32% in metabolism, 13% general function prediction only and 15% poorly characterized (Fig. S8). When mapped to the Archaeal arCOGs using a similar voting heuristic as used in COG assignment, there was a 1:1 relationship for 754 ortholog clusters, 104 clusters mapped 2:1 to 52 arCOGs, with the remaining 36 mapped at higher degrees (24 in 3:1, 12 in 4:1), indicating somewhat finer cluster granularity in our work.

So defined, core genes were rarely located on DL secondary replicons. To explore what persistent contribution secondary replicons might make to the gene repertoire of DL haloarchaea, 68 ortholog groups were selected which possessed at least one secondary replicon member from each DL haloarchaea and also a member from tADL. This set was regarded as conserved but potentially non-core gene content. The largest ortholog groups within this set were associated with ISs (ISH3, ISH4 and ISH6 families) with sizes (72, 27 and 19) which were much greater than was typical (mean size = 9, median size = 7). The next most abundant ortholog groups included COG functional groups for transcription (K), replication, recombination and repair (L), defence mechanisms (V), inorganic ion transport (P), intracellular trafficking (U) and cell cycle, division and partitioning (D) (Fig. S9).

Selection - Ka/Ks ratios (ω). The selective pressure of haloarchaeal genes was considered both within and between populations. Single copy ortholog pairs were identified between tADL and *Hl* (within population) and tADL and *H. volcanii* (between populations). Clustalw2 was employed for multiple sequence alignment and ω was estimated by KaKs_calculator using AICc model selection (-m MS). For 843 within population and 842 between population ortholog pairs (mean ω: 0.065 and 0.062 respectively), none were found with $\omega > 1$ and only 8 within population and 7 between population had $\omega > 0.3$. The apparent distributional similarity for within and between population ω values (Kolmogorov-Smirnov D=0.04, p-value=1) suggests that core gene content is under similar purifying selective pressure irrespective of environment.

Whole metagenome assembly. The whole metagenome of DL was assembled using Celera WGS (v6.1), with the recommended component pipeline for 454 Titanium data. An initial assembly was performed with default runtime parameters, with the exception of 3.0% unitigger error rate. Celera WGS estimation of genome size and consequently its effect on the discriminating statistic (a-stat) categorising early contigs as degenerate (repeats) performs poorly on datasets which do not represent a single organism. The tendency to over-estimate genome size leads to an increase in false positive rate of degenerates. To mitigate its effect, genome size was reduced manually by two-fold and the assembly repeated post-overlap stage. This was repeated iteratively until the rate of change in degenerate assignment began to slow. The previous step was then taken as the final result. In the case of the DL metagenome, genome size was reduced 4-fold from the predicted 88 Mb to 22 Mb. From 6,626,699 usable reads, assembly resulted in 16,551 contigs (mean length = 2736 bp), 634 large contigs (> 10 kb) totalling 12 Mb in total length and 15917 small contigs totalling 33 Mb in total length. An additional 61 large

contigs (> 10 kb) classified as degenerate due to high read- depth (mean read-depth = 200) were also included.

For assembled contigs the low complexity of the DL metagenome permits inference of replicon source by cluster analysis in a two-dimensional space composed of GC content and mean read-depth (Fig. S10). Restricted to lengths greater than 15 kb, contigs belonging to each primary replicon of the three abundant isolate genomes were first identified by stringent BLASTN assignment (coverage > 90%, e-value < 10^{-10}). The mean GC and read-depth of each of these labelled clusters is in agreement with values inferred from FR and reference genomes. Model based clustering using Mclust was performed in R with background Poisson noise to compensate for the presence of outliers belonging to no cluster. As tADL contigs are clearly separated, this step was limited to a subspace (20 < read-depth < 100) containing only three clusters (DL31, *Hl* and the unlabelled cluster) (Fig. S11). The unlabelled cluster (center: GC=0.63, RD=38.3) comprises 52 large contigs (> 15 kb) totalling 1.89 Mb in total extent, subsequently referred to as the "tADL-related 5th genome".

tADL is the nearest relative with an average nucleotide identity (ANI) of 0.802 (1.08 Mb aligning) and TUD (tetranucleotide usage deviation) regression of 0.959. Against tADL, NUCMER reports 492 gapped alignments totalling 806,411 bp with an average identity of 85.2%, while CONTIGuator maps 48 of 52 contigs (Fig. S12) or 1.82 Mb in total extent with 28.8% identity. Coverage across the tADL primary replicon correlates inversely with regions identified as putative genomic islands rich in non-core gene content (Fig. 2 and Fig. S7).

Annotation of the "tADL-related 5th genome" using SHAP predicted 1889 putative full length genes. Genomic content was inferred by orthologous group assignment, where each gene was scanned against the orthologous group profile HMM library using hmmpfam (e-value < 10^{-10}). This classified 873 genes (46%) as core gene content, covering 76% of orthologous groups, while 40 genes (2%) were assigned as non-core gene content. The proportion of core to non-core gene content and substantial coverage of the tADL primary replicon is highly suggestive that these contigs are likely to be of primary replicon origin.

Similarity comparison. ANI and TUD regression coefficients were determined between all replicons with JSpecies along with the "tADL-related 5th genome" for comparison (Table S8). Genomes with similarity scores of greater than 0.96 ANI or 0.98 TUD can be considered the same species.

The spread in similarity score distribution across the primary replicons is roughly 7-fold greater for TUD regression (0.830 ± 0.065) than ANI (0.721 ± 0.0084). Calculation of ANI relies upon BLAST alignments which may cover as little as 20% of the full replicon length, while TUD regression is inclusive of the whole genome and is independent of any preliminary homology search, suggesting perhaps TUD regression has a greater sensitivity in this case.

Primary replicons for DL1 and DL31 show the greatest similarity in TUD regression coefficient (0.945) but are otherwise unremarkable in ANI (0.715). The "tADL-related 5th genome" is most similar to tADL (ANI=0.808, TUD=0.959) but would not be considered the same species by conventional definition.

High identity regions. Though the DL haloarchaea genomes typically possess ~80% ANI, there exist many regions longer than 5 kb of much higher inter-replicon sequence conservation (>99%). Experimental validation by PCR amplification and sequencing confirmed the presence of HIR in their respective genomes (see **PCR and DNA sequencing confirmation of HIR** below).

Thirty regions longer than 5 kb are shared across seven of the nine isolate replicons (DL1: Contig37 and Contig38, DL31: Contig114 and Contig115, *Hl*: NC_012028, NC_012030 and tADL: Contig32) and thirteen regions longer than 10 kb are shared between six (DL1: Contig37 and Contig38, DL31: Contig114 and Contig115, *Hl*: NC_012028 and tADL: Contig32). BLASTN search of regions longer than 10 kb against Refseq_genomic showed that only DL haloarchaea possess sequence identity above 95%, with resultant mean query length 5033 bp. Significant hits below the 95% identity threshold approximated a normal distribution on identity (μ=86.3%, σ=3.4%), possessed mean query length 1317 bp and were associated with haloarchaeal mobile elements.

Represented as a network with replicons as nodes and edges weighted by summation of region lengths over 5 kb (Fig. 3b), the most significant node is for replicon *Hl* NC_012028; ranked first both by normalized weighted degree (0.374) and normalized betweenness centrality (0.567). The top three most significant edges (normalized weights: 0.323, 0.182, 0.161) are intergenic links from *Hl* NC_012028 to DL31 Contig114, DL1 Contig38 and Contig37. From a genomic perspective, *Hl* and DL1 share regions of conserved sequence (>5 kb) with all three other lake isolates (degree 3), while DL31 and tADL share regions with only two other genomes (degree 2).

The identification of HIR raised the question of how novel the finding was, and what would be observed across the many sequenced haloarchaeal species and hypersaline haloarchaea-enriched environments for which metagenomic data exists. To answer this question, an all-vs-all analysis was carried out between 25 finished HA genomes (Table S9). ANI by BLAST was determined using JSpecies and the total extent in base-pairs of long shared HIR was determined using NUCMER. For HIR, only alignments greater than 99% identity and longer than 2000bp were considered, where the length criteria was chosen to minimise bias attributable to short, possibly well conserved mobile elements such as ISs. The total length of the resulting alignments were then summed for each genome-pair. Two symmetric "comparison" matrices of genome-pair values for ANI and L_{HIR} were constructed. Due to the large value range and presence of zeros, L_{HIR} matrix elements were transformed by $x'_{ij} = log10(x_{ij} + 1)$ prior to further steps.

Matrix element definitions for each of the two matrices:

$$ANI_{i,j} = ANI(genome_i, genome_j)$$

$$L_{HIR_{i,j}} = log10\left(\sum_{i=1}^{n} length_{i,j,n} + 1\right)$$

Heatmaps (Fig. 4 and Fig. S14) were produced for both ANI and L_{HIR}, where hierarchical clustering of Euclidean distances was used to reorder rows and columns for minimum variance (Wards method).

Excluding self-self comparisons along the diagonal, the median global ANI of 71.3% correlates well with the primary mode within their distribution (Fig. S14B, red line). Prominent block clusters within the ANI exist, where more closely related genomes occur adjacently: H.salinarum and NRC-1 (>99%), H.walsbyi and H.walsbyi_C23 (98.7%), H.hispanica and H.marismortui (90.6%), H.volcanii and H.mediterranei (82.4%). A larger block of seven genomes also exhibits collective similarity: Nat_J7-2, H.turkmenica, H.xanaduensis, N.pellirubrum, N.occultus, N.gregoryi, N.magadii (median ANI 78.5%) producing an observable peak with the distribution of ANI (Fig. S14B, blue line). Comparisons between the two H.walsbyi species and the other 23 genomes show that the H.walsbyi are systematically more distantly related (median ANI 64.8%) (Fig. S14B, green line). For the DL isolate genomes (tADL, DL31, H.lacusprofundi and DL1) the median ANI 73.1 % (Fig. S14B, light blue line). Fitting a normal model to the primary mode within the ANI distribution by expectation maximisation (μ=71.2, σ=1.4), only the DL isolates ANI can be considered not significantly different from the mean (p-value = 0.09).

The DL isolates between each other exhibit a median ANI not significantly different from the global median between all 25 haloarchaeal genomes, while at the same time there exist other genome-pairs with ANI significantly different than the global median and closer to unity. Not all of these more closely related genome-pairs possess HIR, while all six possible genome-pair combinations between the four DL isolates do, suggesting that HIR are a peculiar feature of the DL community -- not observed in other haloarchaea whose complete genomes are currently available.

For the L_{HIR} heatmap (Fig. 4 and Fig. S14C), the majority of matrix elements are zero as there exist no alignments of sufficient identity and length between genome pairs. Of the 9 genome-pairs which contain HIR regions, 3 pairs are from closely related organisms (>90% ANI): H.walsbyi and H.walsbyi_C23, H.salinarum and NRC-1, H.hispanica and H.marismortui. The remain 6 genome-pairs possessing HIR regions involve only DL isolates and represent the full complement of possible pair-wise combinations. The total length of HIR shared between the four DL genomes ranged from 6,561 to 138,307 bp (Table S10), with the extent of matches between genomes mirroring the results from network analyses (Fig. 3B).

To assess whether the DL-specific HIR were present in haloarchaea from other hypersaline environments, FR was performed independently for each long (>10 kb) DL HIR against 15 saltern metagenomes (11 Chula Bay, 4 Santa Pola; 2.8 million reads, 1.1 Gb) obtained from the Sequence Read Archive (SRA). Recruitment was performed using FR-hit imposing a minimum read coverage of 50% and relaxed minimum identity of 70%. Read coverage across individual HIR was visualised using in-house scripts in R and compared against published gene annotations.

Metagenome sample environments with lower salinity, Chula Bay low (6-8%) and medium (12-14%) salinity ponds, did not recruit any reads. Of 5700 recruited reads, all originated from higher salinity environments (19-37%), Chula Bay high (28-30%) and Santa Pola SS19 (19%) and SS37 ponds (37%), consistent with salinity ranges where haloarchaea begin to dominate the community. Despite low read assignment stringency, coverage across individual HIR was incomplete, with reads mapping primarily within IS associated genes (Fig. S15). Only 1.3% of recruited reads (74 reads) exceeded 95% identity, and none exceeded 99% identity.

To examine the genomic content of HIR (>5 kb) all 654 genes completely contained within them were considered. Assigned to orthologous groups, only 6 (< 1%) genes were identified as core gene content while 180 (27.5%) were considered non-core gene content. By ArCOG assignment 380 (58.1%) were poorly characterised, 177 (27.1%) were information storage and processing, 60 (9.2%) were cellular processes and signaling and 37 (5.7%) were metabolism. Nearly two thirds (63%) of those assigned to cellular processess were associated with V, defense mechanisms.

ArCOGs from [V] assigned to genes from HIR:

- arCOG00719 PIN domain containing protein
- arCOG01208 Minimal nucleotidyltransferase
- arCOG01664 Cytotoxic translational repressor of toxin-antitoxin stability system
- arCOG02777 Restriction endonuclease
- arCOG03779 GTPase subunit of restriction endonuclease
- arCOG03899 Predicted restriction endonuclease, HNH family
- arCOG04493 PIN domain containing protein
- arCOG04793 Transcriptional regulator, predicted component of viral defense system
- arCOG05102 McrBC 5-methylcytosine restriction system component
- arCOG08906 Predicted antitoxins containing the HTH domain

Sequence flanking each region was compared against Refseq_protein with BLASTX (e-value < 10^{-5}), where the analysed regions were 2000 bp upstream of the ends of each region. Of the 25 HIR considered (50 flanking regions), 24 (48%) were associated with transposase genes, an additional 11 (22%) associated with integrase, resolvase or endonucleases, 5 hypotheticals and 10 genes of various annotation.

Insertion sequences. ISsaga was used to analyse and manually annotate DL haloarchaea genomes for ISs. Of the 489 IS related ORFs found across the four isolates, 297 were putative complete, 108 putative partial and 84 uncategorized. Collectively as a set, ISsaga estimates between 127 to 178 different IS types. Three IS families (ISH3, IS200/IS605 and IS5) comprise 65% of all predictions (Table S5). Considering only predictions with greater than 90% similarity to known ISs, a negative correlation was found between the proportion IS related base pairs and replicon length (Spearman p-value = 0.042).

A bipartite graph of IS to host genome helped to visualise the prevalence of each within the community, where nodes and edge size is relative to the number of ISs found within the target genome (Fig. 3a). The node out-degree profile of IS nodes follows a power-law distribution (R^2=0.970) with the top three nodes ISHla1 (38.4%), ISHla6 (20.7%) and ISHla7 (14.6%) possessing 73.7% of all edge weight. Although model fitting to a limited number of nodes is nonsensical, it can be seen that tADL (25.8%), DL31 (31.1%) and *Hl* (32.3%) possess roughly equal numbers of the most prevalent ISs, while DL1 possesses fewer (10.7%).

It has been argued that as the set of nonessential genes decreases with genome size, the frequency of highly deleterious transposition targets acts to control IS abundance (6). In light of this, the high density of ISs observed within the secondary replicons of the DL

haloarchaea suggests that these replicons possess a high degree of freedom for genomic plasticity and are likely to contain predominantly nonessential genes. There exist also regions of high IS density within the primary replicons, in particular a 400 kb region within tADL (1.150-1.550 Mb) which itself does not possess a secondary replicon.

Using the ISsaga database a total of 489 matches to ISs were identified across the four genomes, with three families (ISH3, IS200/IS605 and IS5) comprising 65% of all predictions (Table S5). The density of ISs was noticeably higher in secondary replicons and in distinct regions of the tADL and DL1 primary replicons which possessed low CAI/CBI indexes. The secondary replicon with the lowest density was >4 times the density of any primary replicon, and the density in tADL was >2.5 times higher than the next dense primary replicon. A comparison to 119 other species showed tADL is ranked 2nd in number of ISs behind *Sulfolobus solfataricus P2* which has considerably more than any other in the ISsaga database.

Lake viruses. Accessions for 55 completed archaeal virus genomes were sourced from the European Nucleotide Archive (ENA) (URL: http://www.ebi.ac.uk/genomes/archaealvirus.html) and retrieved from the ENA Sequence Version Archive (URL: http://www.ebi.ac.uk/cgi-bin/sva/sva.pl?&do_batch=1) as DNA fasta sequence. An archaeal virus database was generated from the 55 genomes, where BLASTN identified 15 contigs over 1 kb and 4 contigs over 10 kb with significant similarity (e-values < 10^{-5}) to viruses associated with halophilic *Euryarchaeota* hosts (7). Identifications were comprised of *Myoviridae* HF1, HF2 and PhiCh1; pleolipoviruses (8) HGPV1, HRPV2, HRPV3 and HRPV6; and *Siphoviridae* BJ1. In total extent, assembly contigs associated with *Myoviridae* viruses constituted 68 kb with reported genome sizes 75 kb. For *Siphoviridae* BJ1 with a reported genome size of 42 kb, two associated contigs (29 kb, 53 kb) may represent two separate genomic scaffolds.

Genome characteristics of the four DL haloarchaea. The genomic features that distinguish tADL from the other three DL haloarchaea and may therefore contribute to its dominance include: 1) a single replicon; 2) a physiology that includes a preference for high-affinity uptake of carbohydrates, complemented by photoheterotrophy and carbon storage. Only tADL possesses genes for gas vesicles, bacteriorhodopsin, and polyhydroxyalkanoate (PHA) biosynthesis, and it has a higher number of predicted ATP-binding cassette transporters for carbohydrates (six), and possesses multiple glycerol kinase orthologs (first step in glycerol breakdown) and a large number of regulatory genes (*e.g.* signal transduction). Thus, tADL appears to have a highly saccharolytic (carbohydrate degrading) “high energy” metabolism that can respond to changing substrate availability, with glycerol as a preferred substrate. Gas vesicles provide buoyancy that facilitates upward motion, and particularly for slow-growing organisms can allow more efficient vertical migrations than swimming by flagella (9). This may facilitate tADL getting to the surface in the summer, thereby allowing light-driven bacteriorhodopsin to generate energy, and faster growth rates to occur in the warmer water. This reasoning is consistent with tADL abundance being somewhat lower in the deepest point (36 m) of the lake (Fig. S5 and S6). Additionally, surplus carbon and energy could be stored as PHA, and mobilised for biomass production when other limiting substrates become available.

The other genomes each have specific characteristics indicative of niche adaptation. The DL31 genome is characterized by pathways for protein and peptide uptake and breakdown (including many predicted secreted proteases), and is the only one to lack detectable flagella and ammonia transporters. It appears to be orientated towards proteolytic (protein-degrading) metabolism targeting particulate matter rich in protein, with proteins providing both carbon and nitrogen. DL1 is enriched in genes associated with amino acid breakdown, and is the only one to lack genes for glycerol breakdown. It therefore appears to prefer free amino acids, and be unable to benefit from glycerol, which can be an abundant carbohydrate released from algae in hypersaline systems (10). *Hl* has the most versatile metabolism, with genomic potential for utilizing both carbohydrates and proteins as growth substrates. However, it has comparatively few over- or under-represented COGs associated with metabolism, which may indicate that it prefers to target a broad range of substrates rather than having a specialised metabolism.

Precedent for Antarctic ecosystem distinctiveness. A precedent for Antarctic ecosystem distinctiveness is known for Ace Lake, a marine-derived, non-hypersaline, meromictic system that is located ~15 km from DL (11). In Ace Lake, green sulfur bacteria grow at the oxycline in the lake, appearing in a distinct, dense zone about 1 m thick. The population is essentially clonal, represented by a single dominant member, *C*-Ace (12,13). The physico-chemical properties at the oxycline differ greatly from the upper oxic and lower anoxic zones, imposing a range of specific selection pressures (*e.g.* light penetration) on the community inhabiting this zone. *C*-Ace is predicted to have evolved dominance through mechanisms allowing phage evasion linked to a growth response controlled by the annual polar light cycle (13). The whole water mass of DL provides different, but similarly constraining conditions, particularly the specific combinations of extreme hypersalinity and cold; conditions that have selected for an overall haloarchaeal community composition that differs greatly from other salterns in the world where *Haloarcula* spp., *Hfx. volcanii*, *Haloquadratum walsbyi* and *Halobacterium salinarum* are typically found (but are absent in DL) (14-16). By restricting the nature of species that can grow and compete in the lake, and providing conditions that naturally promote gene exchange, the system appears to have had the time to enable gene exchange events to become relatively frequent and fixed in the population, becoming an important driver of haloarchaeal community evolution.

PCR and DNA sequencing confirmation of HIR. PCR and Sanger sequencing was used to confirm that presence of HIR in each of the four DL haloarchaeal genomes, by amplifying and sequencing the boundary regions of randomly selected HIR from all four strains. DNA was extracted from cultures using the xanthogenate-sodium dodecyl sulfate extraction protocol. Primers were designed to ensure that the desired amplicons would include sequences immediately before and after the boundary of the HIR (Table S11). The following is Sanger DNA sequence data obtained for PCR amplified fragments, noting the organism, replicon and genome location.

DL1

Replicon: HalDL1_Contig37 Shared region: 20871..25836

CTAAACACACACCTGGGTATCGAAGAACCCATCTTTATTTCCGATCAAGGGAGCCCTGACCTCCTCAGTCCA
AGTGGTCAATCTCCGCAAACAGATACCCCTGAAGAGCGCTGACTGTTTCTTCTGCGACAACTGAGGCCCCGA
AGTCACTCCGAAATCCATGCTTGAGTTCGTCGCTCTCCAAGATTTCCAGACCACGCTCAAGCTGCTCCCGGA
GCGCATCGTCGTCCCCTCTCCGCAGGTGAGGCGTGACAGCGTCAAGGTCGATTTCTTCTGCCACTGCTAGGA
CATCCCGGAGGTCTGTTTCACGGCCGCTGTGGAGCTTTGCCGCCACGAGGACTGCCCCATCGATGACTCTGG
CTGTGGTCCTCACTGTACCTCCGCTCACCTCCTGTTGGTGGCTGTGGTCGTACAGGTAGTCAAACGACCACT
GTGCCTCCGTCTGGCGACACCCGAGTCCGTTTACCAGGAGATCGAAGCCGATCGGCTGCTGCGGCGTGAGCC
GCTTCTCGTACTCGATTACCTCGGTATCGTAAAACCACTCCTTGGCGTGGCTGTCCGTTTCTTCGAAGCCCC
GCTGTTCGAGGAACTCGACGAAGTCAGCCTTGGAGTCTGGCGCGACGACAATATCGAGATCCGTGGAGAAGC
GAGCATTGAACGCTGAGACAGCGTAGCCGCCAACAAGAACGTACTCGTGACCCTGTTGGGTGAGCTCTTCGA
GCAGTTCGATGAGCGCGTCACTTCGATTGTTGAAGCTCATGGCTGTGCCCGTTCGGTCTCTCGATACGTGAC
GCCGAGGTCGAGGTCCTCGTACATCCGGTCGAGCATCGCCAGCCCCGACTGAAATTGGGCGTAGTTCTCGCG
CATATACTCGATTGTCTCTGCTCTCGGGATCACCGGGTACCCTTCGACGTGCTCGATGTCGAGTGACGCGCG
TGGCTCGAGGACGATCTGCAGCGGTCCGTCCAACTCGTCTTGGGGCTGTCGTTCGATGCGGTGGGGAGGTCG
ACGACTCGAAATGTCTCCAGGCGTCGACGTCTGCTCACGACAGCGAGGACATGATAGTCGTCGGCTCGCGAC
GACTGTAGCCGCCTGATCCACACGTAAACGGCGTCGATCCGCGGTGACGCGACCGCAGTCATGGACTGTGGA
TGGCTACGTACCTCGATGAGGTGAACTGACCTGCTTGCAGCCAAGTCGACTGATCGTAGGCCTCCACACGTG
GAGCGCTCTCTAGGTGAGCTACCGATTCG

TCGCGGCCTTGGGAAGAAACCGAAGGAGTTTGCCGCGGCAGCCGAGGACGATACCAGCCAATATCTTGTTTG
AATGGCGACTGAGGAATCGCAATATCCGATTCTCGTCGATACTGACGCCCTGATCGCTGTCGGGAACAGCTC
ACTATGGGATTGTATTACTGAGAACATCGGACTCACGACCACGAACGTCTGCCAGCAGGAGCTAAAGCGCCA
TTACGAGCAGAACCGCTCTCACGCTCCAGAAGGAAGTCGCTCGTATCGCCTTCACCACGGTAGTACGCGTGT
TCTTGATGCGCTTGAGGATACAGAGACACCGTTGACACGTGTCGTGAGTGTCCCTCGACCGCATGGGGCTGA
TGCCGGTGAACAATCGCTGGAACAGCATCTCGTGCAGCATCCAAAGGCGGTCGATTACGTGGTGTTGATGGA
CGCTCACGGCCGCCGTTCGATCCGTCGAGAGATTCAGAACCGCGACCTCGAGGCGCGTGTAGTGCCACCGAC
TTTTCTCTTCTACATTCTCTACGATAACGATCTCATCTCGCGAACAGCGTTTTGTGAAACGTGTGTTGAACT
TCTCAAAAGCGAAGGGTGGACAGGCTATCACGCCGTCGATGCTGCGTGGGAGGGGATTCCAGTCGACTGCTC
AGACGTCATTGACTCGGATCTCCTTCCACCTTCATGATTATAAGGCTCTGTTGAAATCCCATTCTTCAGATA
GATTCTCACCTGAAACAGCCGGTCGCTGAAGACTGCGAGTTGACTGATTCAAAATCGCGGTGGTACTGTTGG
TTCCGTTGTTCTCGGTGAGCCTGTGGCTTGACATCAGGGCAGTTTGACGGAGAGACTGTAACTGGTCTCCTA
ATCGGATTGTTTACGGTGGCGTTGCGCTCTTACATCTGGTTTCGATGGGCGTCCCAACGGCCCAGTTGCTGA
CGGTACCCGCAAGTATTTGGTACATATACACACGTTACAGGCGCATCGGGCCTTCGTTGAAATAAATACAAA
CGCTCGGCTCTGTTGAAATACTTTTTCTCAGACATATACTTGCCTGAAACGGCTGGTCGATGGAGACCTGCG
TATGACCATCAGAATCTACAGCTCAGAGGGATGTTTCGATAACGTGCGAGAATACAGGCGCAAATCGGTTAC
GACTACTCTAACCATCGAATGTCTGGTGTGTCTGATGAATTAACAATACTCCTCAACTCAGGATGAAGAGAC
GTTACATGAAGTGCTCCGCAACGCGTTTGCCACTCTCTCTATGTAGACG

Replicon: HalDL1_Contig37 Shared region: 44074...49706

TGGGGGGCACGTCCATTTGTATAAGGACCGCGAGAAAGAATGTTCCACGCTGCGTTCGCGTCCCTGTCCGCC
TGAAACCCACACGAGGGACACGAGTGTTCACGAACCCACAACGGCTTGTCCGTCTTGACGCCGCAAGACGCG
CACGCTTTCGTCGTGTTCCTTGGATCGACTTCGGCAAAGTGTGTTCCTTCGCGTTCGCACTTGTACTCAAGC
ATTCGAAGGAACGTTCCCCACGCCGCTCCCGCCCGATTCCGTGAGTTTCCGGGGAGTTCGACCAACTCCTTC
GCGTCGAGGCCTTCGACCGCTACGAGGTCGTATTCGGTGGCGTAGTAGTTCGAGAGTTTGTGAAGGAAGTCA
CGACGCTTTCGCTTGAGATCGGCGTGGCGCTCGGCCACAACTTGCCGTTGTCTCTCCCAATTCGCGGAACCG
CGTTCCTTCTGCGAAAGATCATGCTGTGCGCGTTCCAACCGCTCGCGTTCGTCAGATAAGTCGAGCGATTCG
ACGGCGGTGCCGTCTGTGTCATGGGCGTACTTCAAGATGCCAACGTCAATCCCGACACATCGATCTGGGTCG
GTCGGTTTCTCAGGAGTCTCTTTATCGCCGACGGTGAATGAAGCGAACCATTCACCGGTTGACTCCTTCTTA
ATCGTGATTTCTTTGACAGTCTCGGCGTCAGGAATTGCGCGGTGTTTGATGAGGGGAATATCCGCAAGTTTC
GACAGCGACAGTACGGGTTGGCCGTTCTTACTATCGAACTCGAAGCCAGACTGTATGTAGGTGAAACTACGG
AAATCCTTGGGGGCCTTCCAATTGAGACTGCCCACGTTGTAGCCGTTCTGCTTCAACTGAGAGAGGGCTTTG
ATGCTGTCTTCGATACGCATGACAGCAGCTTGTGCGACCGTTGAATAGACATCGTTCAGCTCATCCCACCAG
TCTTTGAGGCTGGTGAGCTGATCGCGTACTTGTCGCACTCGTTGGTTAAGTGTACCCGCCGATTTGGGAATT
TGCTTGAATTCGTTGAGTGCGGTGATTGTAGAGTTGCCTACAGGTATCTCGGTGATGATCCAACAGTTCACG

CTGTTCAGCGTGGGATCAAGCCGAAATCTGTAGGTGTAGTGCATTGGCTATCGTCGGGAATCAGAACGGTCG
GACGAATTTTCACGTCAGCTCTCGCCCACCGTTTCGGACCGAGAAACGGTGGTGCGCGATCGGTCGGCCTCC
ATTCCCTCGATGACTTCTGCCTCAATCATGCCATCATCACTGTTGTTATAACTTATAATCAACCTC

CCTCGCTGGATATAGACGAACAGCGGCATTATGCCCATCGACACCTGCGCGAAGCCGAACCCGGCGATACCG
TGAACGGCGGCCGCGGCGAGGATAACGACCGCAGACAGCCCCGTCGTTACCCAGTCGCTCGTCATCGATCGC
GCCTCCGTTGCCACGGTAGAGTACGCATCTGGATGAACGATACTCCGGCTGCATCTTTTATTGTAATGCTCA
ACTCTGAAGCAGTAAGTGAATCATATATTATATTTACTATATAAATAATTATAACCGGATATTCCGAACATT
CGGTGTCTGTAACGACAGTCACATCACGATGAACGAGCATGAAACGGATATCGGCGAGTGGCGGTCCCTCGG
CGGCGCCCCTGTCGGAAACGGTAACGATGACGCCGACGCCGTCTTCGCTCACGTCAGCGACCTCCACGGGCA
GCTGACGCCGCGCTACCAGGTCTACTACGACAATCCGACGTCGACGCCGGACTTTAATTTCGGAGACGACGA
TCGCGTCGTCGAGCGCGGCGGCGGGATCCCCCTGCTCGCAGCGAAACTCGACGAACTCCGCGAGGACTACGA
CGTGTGTACGCTCATGAGCGGCGACACGTTCCACGGCTCCGCCGTGACCACCTACACCGATGGGCGAGCGAT
GCTCGATCCCGTCAACGACCACGTCGCGCCCGACATCTATGTCCCGGGGAACTGGGACTACTCGAACGAGGC
CGCCGAGGACGGCAACTTCGTGGAGTTGATGGACGACCTCGACGCCCCGATTCTCGCGAACAACCTCTACGA
CTGGGAGACCGACGAGCGACTGTACGACGCGTACCGGATCTCTCGACATCGGCGGACTCTCCGTGGGGGTCG
TCGGGATGACGAATGTCTACGTCGATCGGATGGCACCCGCGTTCTCCGAGGGGGAGGTACCGCTTTCGGTAA
ACATCCTCACACTTCTTCGAGGAGTCCGCACATGGCCGCCCGCCGAGGACGGGCGCGGACGTCGTGGTCGCG
GTACCGAGATCGGCCTCCCGTGGATGGTCCAAGCCGCCAGGACTGTGCGAGCGTGGACGTGATTGTTCGTGT
CGCCACCTCCACGAAGTACCCTACGAATCGAATCGTCGTCCAG

Replicon: HalDL1_Contig37 Shared region: 49699..61930

GCTTGATCCAGCTCCGGCCGCCATGGCGGCCGCGGGAATTCGATTGTAGAGTTCGCCGAGCGTGATGTCGCC
GGGTGGGATGGCAGTCCCGTACCGGAACCCGTGCGAGACGGCGAGGTCGGTGCCGAAGTGTGCACGGAGTGC
GTCGTTGAACAGCGCGTTCCACGCGCTCTCAAGGAAGGACTGCCGGTAGAGCGGTTCTTCCGTCCGACCGAC
GACCGCATCCAGCGGACGGTCGAGCGTGCCAGCCCCTCGCTCGAATCCCGGATCGGCCTCGAAGAAGGGCGC
ACGCACGGCTTCGACTGTCTCCGCCGCATCGGCGTCCGGTTCCGGCGTGTGCTCGCCGTCCTCGGTCAGACA
GTAGAGGTGGTGACGGAACTGTATCTCCCCGTCCCGAACGCGGAGGTCCACACGGCCGATCGCCTCACCCAT
CCCGGACTCGACGACCACGGTTTCGGTCTCCTCGACGACGATCGGATCGTAGGTGTACTCGTGGGTGTGCGC
GCTGAACATCACGTCCACGCTCGCACAGTCCTTGGCGGCTTGGACCATCCACGGGAGGCCGATCTCGGTGAC
CGCGACCACGACGTCCGCGCCGTCCTCGCGGGCGGCCTGTGCGGACTCCTCGAGGAGTGTGGGGTGTTTACC
GAAGCGGTACTTCCCCTCGGAGAACGCGGGTGCCATCCGATCGACGTAGACGTTCGTCATCCCGACGACCCC
CCACGGGAGAGTCCGCCGATGTCGAGGATCCGGGTACGCGTCGTACAGTCGCTCGTCGGTCTCCCAGTCGTA
GAGGTTGATCGCGAGAATCGGGGCGTCAAGGTCGTCCATCAACTCCACGAAGTTGCCGTCCTCGCCGGCCTC
GTTCGAGTAGTCCCAGTTCCCCCGGCACATAAGATGTCGGGCGCCGACGTGGTCGTTGACGGGATCGAAGAT
TCGCTCCGCCCTTCGGTGTTAGTGG

Replicon: HalDL1_Contig37 Shared region: 70308..74648

CTGCTCCCGGCCGCCATGGCGGCCGCGGGAATTCGATTGGGACCTCAATCAACCACGTCTGCGAAACAACTG
ACGACTCGCCCCACGCCAATACCGTCCGTGGACATCTCACCGACCAGTTTGAGCTGGACTCCGTTGAGGCGG
TTGGGGACACACTCCTGCAACGAGATACTCTTGAGACACTGCCGGATCGGCCGGTGGAGGTCGTCGCCTACC
TCCAGCTGGATCCTTACTACGGTGACGAGGACGAGAGAGATGCGCTGTACTCCTCGCAGGCTAAACGCGGAA
CCATGTCCTTTCACGCGTATGCGATGCTCTATGCGCTGGTACAATCGTGACTTTATTACTGACCTCATCATC
TAATCGGATGAGTTGCTCTATGGACGGTGTATCCGCGAAGATTTCTGAATGGCATTCCTGCAATAAGTCGTA
CTCCTAAGCAACCTATCCACCCATTCGTCTCACCTATTCTCTATCAGTCGAATTAATGTTTTCGTGGGGATC
CCCCCCTGTCGTCTCCCGGGACCCCGACCCATAACGACTTCCATCAGGGGTGGTAATCGACTGATCCCTCGT
TACGTTCTCACAACCGATTAGCTAGTGACCTATCATCGGGCTATCACATCGAGAATTCTCTACCCCCTCCAG
TAAAGGTTGAATTTGCGTCTTCGATGCGTGGTCAATTCACGAAATGACACGGACGGTGAGTTTCGGCGGTGA
GGGGTCGCTGTCGGCGGC

CCTAGTCCAGCTCCGGCCGCCATGGCGGCCGCGGGAATTCGATTCTTGTTCCAGAAGCGTCGTGGCCTCAAC
GACTCGCTCTCGGTCGGGAGAGTCCAGTTGCTCGATATCGATGCTGTTGAGGTTGCTGAGTGCGCGATGCAG
TCGGTATCCTGGTGTGTCTCGCGGGTCCATCATGTACGCCTCCGCCGACTTCTGGCGCAGAAAAACGAAGCA

GGTACGCTGATGTCTCCCGCTGAAGCCCTGCACCAGCCTTAACCAACAGACTTCGTGTAGGCATCCTGTGTT
GGTTAAGGTGTTGTTCAACCGATTATCCTACTCGGGGTCTGGTCCGTGACGAGGTGACTGACACACCCGGCA
CTCCCACATTGGCTGGCCAGACCATTCGTCATCTGGGAGCGGCTCGAACTCCTGGAACTGATGTTCAGTCGT
TCGACCGCACTTGCGACATTCAAGACGTGTTTTCGAGGGAGCGTTGGGTTGACGATTCTCGGACTGGTCGTC
GTCGACGGAACGCTGAAGTGCAATCGCTGGCACTAACCAGACGTGTTAAGTTGCGAATGACACAGAATGTTC
ACGACCTATTTTCACAAGAAATCGCTAAGATATGATCCCTCAAACATGAGTACTCATCAATATTTATCTAGT
ATACCAATCCAAAATAGCCAGTTTAAGGTACATAGGCTTTGTTGAAATCCTATTGAACTAACACTTTTAGCA
GTGACACAGCTGCTACGTAGATGTGGGAATTGGCTGTTTCTTGAGCGATTACTCACTTACAACCCGTTTTTT
ATAGACTAGATATGCTAATGGAAGCATTAGAACATAGATGACGGTAGTCATGGCTAAAAATAGCTCACTAGA
ATCTACTGCTAAACTGTAGAATAGACCGAGTCCAACGAGACCATATGCTACAGCAAGTGTCCAAATTACTTT
TGACATACATAATGGGTTAATGCCTGATGAGATAACGATTGTGGGACCTTTGCTTCTGTACTGATCTTAGAG
GGCTCGGTTCGATTATTGTGGACAGACATCTAGACGAGTTCACGGAACTTGTTTTCCACCATTGTCGGGAGC
GACGAGCACCGTACTTTCGTTTGGACGTTGAGTTGACGTTCTTCTGATTGACTTCTTCTGTTCGTAGAAGAT
CAGCGTCTAAGCGTGGCGTTTCCATGCCCTTGTTG

Replicon: HalDL1_Contig37 Shared region: 101134..105279

TAAGTGCTGCTCCGGCCGCCATGGCGGCCGCGGGAATTCGATTGTTTCGCAACCAGATACGCTACCGGTAAC
ATAAGTAAAAAGATTAGTCCCAACGCGTAATTCAAGATTTGATCTGAGTCTGTTGACACACCATATAGTAAA
ATGATGAGTACAATCGCATGGAGGGCTGCTAATACCCAGATATACTGATTTTGTTTCATATTGATGTAGATG
ACCAGAAATTATATAATTGTTCGGTTAGGTCCCACAATCAAGACACTGGCCACAAAAGACCGGGAGAGTAAT
CCCACAAACACTCGTGAAACAGATAATAAACGCTACTGCGCCCGGGACGCCAAAAGTTAAGAAAGCCAAACA
TGCTGGACAGGCTCCTCCAATGATTCCGATACACGTAGTACAGGTATTCAAACAAGACTGTTGTGCTATAGC
TACTTCTCCAACTTCAATATCATCCGAGCTCGGTTCGTCGTCTTCGAATAATTCTTGTACTTCGTACTGAAT
CTCACGATCCTCTAAGAGCACCTCATACACTTGGAGATCCAAGTCCGAATCTGGATCCTGTTGAGATGATTG
TTCTGACGAATCTTCGAAGCCTTTGAAGAGGCGAAATCCGTTCATTTCTGAGGTTTTGACAATCCTCAAATC
TTCTGGCTCTGTTGAAACCCTATTCTTCGTACATATCCTTGCTTGAAACGCCCCAGTTAATGAGAGCTGCGT
ATTGACCAATAAAAAGTCATATAATACCCATAGCGTTTTGTGGGTGTCTGGTGAATTACATGTATGAGTGAA
GAAGAACTCGATCCGGTCGAACAACTTCGTCGGGTCGGGATTGGACTGGTACTCGGTGGGCTGGCCTTTGGT
GGGCTGTCGTTTGGTGTCGATGCGGTTGTGAGCGGTATCGTGCTGCTGGGTGCTGGTATCATCGTCATCACT
AGTGATTCGCGGCCGCCTGCAGGTCGACATATGGGAGAGCTCCCAACGCGTTGATGCATAGCTGGATATTTC
TATAGTGTCACTAATAGCTGCGTAATCATGTCATAGCTGTTTTCTGGGTGATTGTTATCCGCTCATCACACA
CATACGAGCCGAAGCTAATGTAGGCTGGGTGCTATGGAGTGACTATCAT

Replicon: HalDL1_Contig37 Shared region: 102306..105542

CCGAAGCAGGGGGCTTTCTTACTTTTGGCGTCCCGGGCGCAGTAGCGTTTATTATCTGTTTCACGAGTGTTT
GTGGGATTACTCTCCCGGTCTTTTGTGGCCAGTGTCTTGATTGTGGGACCTAACCGAACAATTATATAATTT
CTGGTCATCTACATCAATATGAAACAAAATCAGTATATCTGGGTATTAGCAGCCCTCCATGCGATTGTACTC
ATCATTTTACTATATGGTGTGTCAACAAACTCAAATCAAATCTTGAATTACGCGTTGGGACTAATCTTTTTA
CTTATGTTACCGGTAGCGTATCTGGTTGCAAAACGAGCTGATGTGTTATGAATGGCCGCGATTCAAATACAA
CCCTTGATGAGTCAACGCAAATTCAGCGACTTGGCGTCGGTGCCACGGTTGGCGGCCTCATCCTTATTGGTG
TCGTGTCGGTCAGTTCATCGTTTAATGCACCGAGCGTCCTGATTGCGCTTGGCGCATTACCTGTCCTCGGTG
AATGGATAGTCTCCCGATCGTACACCGTCGGAGTTGGAATCGGTATCCTCGGCATATTCCCGCTTGTCGGAA
AAATTCCCGTCAACCTCCGGTTTATCGGCGGTGCACCATTTGTGGCAGGTGTCGTCGTCTACCTGCTCGGGC
CACGCCTCAACACCGTATAAAAATATCGTATGTGTTTACCCCCAAAAATCGCAAGACAGAAAAGGCTTTTGC
TAGCGCCCAGCTTCACCGCTCAACGGCGTAACGTCACTAACACCAGCGCCGTCTGTCGATTTTTGAATTCAC
CGCCAAACAGGATCACTTATAAATGCTATTTTGCCCCTTATTTTCGGGCGTTTCAGGGGATTGAAACCCTCT
CTGTGGGGTAAACAGATACAAAATATCGAAACGACGCCCTCACAATGGTTGCAATCGAAACTCACGACTGAC
GAACAATACGGACTGTCTCGCGTCGACGACTCTCGCTGTCAATCAGAGGTGTCGTCTACGTGTGCTGCGACC
GTGCGCAAACGACATGATGCGATGTGACAGTACGAAACAAACAACAAA

CCGAGGAGAGAAGTGGTGCCCCGAAAACCCGGAAGTCGCCACTATTCTCGGCTTCAGCCGGGTGCCCGACGA
ATCAGCCTTCTCGCGGGCGTGGCGAAATCGATTCGACAACGCCGTTCACGAATACATCCACGCTGCCGCCCA
CTTCGTCATCAAGGAAGTCCACGATCGCGACATTTCAGCGCCCGAGGTTCGGCCAAAGTCAGAGATCCTCAA

CGATACTGAGGAAGCCGCAGACTCAGTAGAAGACGAATCCTTCTCACAGGAGGAGATTGTTCAGACAACGCG
CCTCGCGCGTGATCACGCCTTCGGACACTTCGACTCTGGTCGGGCGTCGAACGCCTCGTACGAGGACACGCA
ATTTTTCGGTCGCGTCGCAAAGCGGTCTAGAGTTTCGAGTTATGGTCATGTAGACTACTCGGCGGTGCTATC
TCTAACGCTCTCTAAGCTACGTGCACTTGGTGCACACCATTCTAACGCAGTATTTTTGTTTGTACTAAGTAC
CGTCGGGACTCTCAGCGGCAAACTCTAGAGGACTTTGCGACACGACCAAATATTTAGATTTCTACATGTTCG
GTAGTTCCAGTACTTCTCAGATCCTGACTTCTCAGATCCTGTGTTATCATTCTATGTTCATATTCAATCTTG
GTCTGTTTTATCAGAGTTCGTTGTTGACTGAGAGATAGTCTGTTCAACTAGTGGTTCGTCCCAGTTGCCCCA
CTCGCCAGGTTGGTCACCCAAACTGACAATGTAATAGAGGAGGGCTTGCCAAAACAAGAAGAGAACGAACCC
GACGGGAAGAGTGAAAAACACCACTTCGAACGTACTCAGAGACCAACGAATGATGATTTGCCACACAACGAG
TGTCACACATGCGCGAGAGACACACTCTAGCGCA

Replicon: HalDL1_Contig38 Shared region: 453313..459149

TACTGTGCTGCTCCGGCCGCCATGGCGGCCGCGGGAATTCGATTCTGATGGTGAAGATGCTGACCGAGTACG
ACATTGGCTCACTCGACGTGCTCATCGGGAACGCCGAAGACTACGCCGACCAGGTGAACGAAGTCTCGACGG
CGAGGTCGCTCGTCCGACTGCGACAAGACGATCGACTCACGATTCGATTGAAGAACCGGAAGACCGTCCACT
CGAAGATTTACCGGATCGTGATGCCGGACGACACGGTGAAGCTCGTCCAAGGGTCAGCGAACCTCTCACGGA
ACTCTTGGGAGTACCACACGAACCAGATCTCCGTCATTACGACCGATGTCGGGACCGAACTGGACGAGGAGT
TCGAGCGGTTCATCGACGAGTACCGCGACGGGTACAGCGACCAGACGCTCCTCGAAGGGCTTGTCGAGGCAC
TGGAGGATGCCGATTCACCGGAAGAACGGGAGAACCGCATCGAGTACTGGGTTGGTGCCGGCGATCTCGACG
TGAGCGATACTGCTGCTCTGAATCAGGACGCCGTTGAGGACCTGAAAGACGTCGCCGATCAGGTCACTGCCG
TTGTCGACGACCCGGAGGGGGCCGACGAGACGGTTACGTTCGTCGAAGAACCTGAGAACGCCGATCGCAACG
TCATTGAGCCGGATGAGCCCGCGGACGAGGAAAACTCCACCGACGCTGCGAACGACGACGAGTCACCGGACG
TCGGGCTCGTCGAGTCGGATCACGAAACGGGGCTGACGGACGGGCTAGATCGTCCGCGGGTCCGTGCACCCG
ACGAGAAGATACGGATGGGAACCAGCAAGGTGGACAAAGACACCGCCGACGAGTTTGGGGCCGGGCTCCGTG
ACCGCGGCGCGACCGTCGAGGATCACAGCATCACCGCGCCGCTGAGCGCGTACAACAACCAGGTCAAGGAAT
CAACAGCCATCCCAACGATGTCCGTCCTTCCGGAAGCCGAGCAGGTCGTGATCGGTGAGGACGACGAGATGA
TTCTCGTCGCCACCAACGAGCCGACCCCTGAAGTTCTGGATCACTGCTCGAAACTATCGAAGACTACATCGA
GTCGTCAGACACGGCACACGCAATCCGAGATCGCAGTGATGGCGCAGATGTACGAGGCATTTCCTCTTACGG
GTCTG

GGGGAAGGCTCTTCCCCATATGGTCGACCTGCAGGCGGCCGCGAATTCACTAGTGATTGCAAGCCCCGACTA
AGACAGTTCATACCAAGGCAGACGGTGCATATAGACTATCCGAATATTACAAGATGTTATCAACGACTTCGC
GTAGGTCATCTGGGCTGGATCAAGCGAAGCATCAGTCGTGGATCTTGAACAGACGGTATTCGTTTATTTTCC
GCGGCAACACGTACGATTTAACGAAATTGTCCGACTAGCGTATCGCTACCAATTGACCGAGCAACCGCTGGG
TCTCGTGTAATACTCTGCAAAAGTTGGACGGGGAATCGAGGAGATTAGTCCCCGAACAGTCGAGAAAAGAAT
CCGCCGGAGTTGTCCGTATCCGATACGTTTTCTGTGCTTGCTGACTGCACGGTATCCGTGCCGGTTTCGGCC
GCTCCGTTCAGGAGCTCGTTGACGGAGTATCCGATCCACTCGGCGAATTCCTCGGGTGCACCGATGTAGACG
CTTGTGGATGATTTTTCGGACATGAATCGTTTGGGGAGGCGCTTCTCGTAGTCGTAGACCTCGTACTCGTCG
CGGTCGAAGTCGGCGGTGATTGCATTCGGCTCAGATTCGAACGTGACACGGCCGCCTTGGATATCCCGCTGC
CACTTCAACCGTCCTGTGTCGTACGTCTTCTCGGCTGGTTCGTCAGCGGGGAAGTACTCGGGTTGCTTTCGA
TCGGCTTCGTCGTAGAACTCGCGGACAATACGGCGGACGTCTTCGATAGTCCCGTTCCCGTCGGTCCATGGT
TGTTCGGTGAGCATTCGGCGGGCTACGTAACTGAAGACCGGGTTTTGTCGTTCCAGAACCTCTGAGAGGTCT
TTTTGTGCTTCGCGGAAGCCCGGGTCTTCCGGGTTGGAGGGGAAGGAGACGTCCATGCTCAGATTTTCATTC
GGTTCCGGAACTCGGATTTCGGGAGCGCGTCGTTGGTCGTGAGATGAGCACGGCTGATCGACGCTGGTGGGT
GTCCAGTCGCCCCAGTAGTTTC

H. lacusprofundi

Replicon: H.lac NC_012028 Shared region: 7615..29467

GACCCCGGGGGAGGGGGGGGGTTGTAGTAATTGAGCGGAGGAGGGACGGTCGAGGCACTTCAAGCGTACAT
CGACGCATCCCCGTGTCTCAGTCTCTGATGAGTACGGTCGTGAGCCCTTGCTCACGACCCGCAACGGAAGAG
TTGCCAGATCAACGATCCGGAAGTCAGTCTATCGGTGGAGTTGCCCACAGGCGCTCGGAGATGAGTGCGATC

ACGATGATTCGATGACATCGTCTGACGCGTGGCGATGTGCAAACAACGCCTGTCCGCACATGGTTCGCTCGG
GTGTGATCACGTATCTCCTCCGAGAAGACGTCCCAATCGACTACGTCTCCGATCGTTGTGATGTCAGTCCAT
CCGTGATTCGGACGCACTACGATGGCCGTGGTGAGTCCGAACGCATGGAAGCCCGAAGTCAGGCAATTGCAG
AGCAGCTCGCTAACTAACCTCGGTTTTCGAATTCATAGACATGACAGAGTTCCATGACGCAGATAGCATCGG
CACCGATAGTGATACGCATATCCTCCCGTCCGACCCAGATATCCACTCCCATGAGCATCAGGAGGTGGATCG
ACGACAACTCCTATGCATCGATGGCCTGACGGCGTACGAAGTCTTCTTAGAATGCCAATCAGAGGATTGCTC
CGAAGAGGCGAGTCTCATCCTCGCAGAAGGGGACTTCCCAGATTCACCGGAAGAAGTAGATGACCTGATGTA
CGACGTGTATCAATGGTTCCGCAAGAACGCTGCAACTCTAGTACCGTATAAGATGGATGCTTCGCTGTATAT
CGCAACCCACGTCCTGATCGCTGGTGAGGAAGATAGTCGCAGTCTGTTACGATCACGGATCGCCCTGTTCCT
TGCTGCCGATGAAGGATATCAGGACGGGGATCACAGTCGCCTCCCGAAAAAGACGGGGAGCTCCTACTATAC
TGTGAAACAACAAACACGGGACGCAATCCAGCTCGTGGGGATCGGGTCAGATCCAGCGCTCAATCAGCTCTA
CCGAGAGATGGGGGAATAAATCATATTTAGCTACGTTCGTCTGACAGGCTCAGTCAAACTCGGTCGCAGGGG
TCGCCAGACAGGGGCGGTGGAGCTTTCTCATGACTCAGTCAGAAGACCGAACGGAAGGTATCGGATAAACAG
CTCAATTCTCGAGAATGCCTTGTCACGCTTCGATCTTTTGGCCCATCGGAATCCCCTCTTCACAGGGGGGG

GGGGGGGACGGTGGGTTTTGCGAGACGACCGGTGAGGTAGTGCGGGAGCTCAGAAGCAGTCCGCTCTTTGGT
TCCACGTGCAGCTTACTAAGTGAGAAGCCAGGATTGATTCAGGACGAGGGTTTATAACTGGTAGGTTCCTTT
GAGCAATCAAGCGACGCCCGAGAATCACCCCGCATCTACAAGCTGACGATGCATAGACCCTCTCTGGCAAAA
TAATTCGATACCGCAATCCCTGGTATCATAGATCCTTGCTGACGCGAGGCCAGCGCGTCGTCGGACGTCGCG
TGCTAGCGGTACGATCAGCAAATAGCACGTGATACCAATGACGTCAGAAATCCCGCCAGAATACGACGATAG
CACAGTTCGGCATTATCGCGCCGACGGTGTCCTCACCGTCACGCATGACGACGAAACAAACGAGCATGTCAT
CAAATCCCGCGGCCGTGATCAGGGTGACAAGTGGACGCGTCGCGTCCCGGCAACCCGGACTACCGTGGACGT
AAGCGAGGCCCTATGGAGCATCCCGGACAACTGGACGAAAGTGTACAGGCTCACAGCTGAGTACGGCCATGA
TAAAGGTATATACCGTATCCCCGAGTCCGGCGACGCGGTCCTCGTGTCGCTATGTCACAAAAACAGCCACAT
CACTGACGCGAATCACACTGTCGAAGCGGTCGGGAGTATAGCCTGGGATGCTCGTGCAGAAGTTGACGAAGA
GGCGCTTGACGATGCTTTCAGCCATCTCAACGGCTGCGCCGATGAGTTCCCCGACGGCGTCCGTGAAGTCCT
CCAGTACATCCGTGACAACCCGCGAGAAGCCGTTGCGGACGCTGAAAAGGACGCTGAAATGCACGCGGTAGA
CTGCGTGACGGATTTGGAGACGATCCCGGTATCCGAGTTTGACGCATTCCGGGTTACCTTCCGGTCCGAGGG
GGGGAGTCGTGAGTCACCCAGAGTACTCATCTGATAGTGGAGGTTATGGAGGCCGTCCGCGAACTCCTCAGC
GAGTCAACGTCGTTCCGCCGTCCCGCTTGTCTCGGTGACAGTCCGTAGACGCGATCGCGTTTATGAGTAGCT
CCGCCACGAGGGTCGAGTGGATACGCGCACTAAATCTGACCACGACTACCCGAGTACCGGTAACCTGCCAGC
TCGAACTCTCTAATCAGTCCTGTTCGACCTTGAACGTCGTAAGACTGCCGCAATCGAGTTGATCGAGCAAT

Replicon: H.lac NC_012028 Shared region: 54359..60782

CTAGGCAATTTTTCGTCGGCCCCTCCGGGTCGTCGACAACGGCAGTGACCTGATCGGCGACGTCTTTCAGGT
CCTCAACGGCGTCCTGATTCAGAGCAGCAGTATCGCTCACGTCGAGATCGCCGGCACCAACCCAGTACTCGA
TGCGGTTCTCCCGTTCTTCCGGTGAATCGGCATCCTCCAGTGCCTCGACAAGCCCTTCGAGGAGCGTCTGGT
CGCTGTACCCGTCGCGGTACTCGTCGATGAACCGCTCGAACTCCTCGTCCAGTTCGGTCCCGACATCGGTCG
TAATGACGGAGATCTGGTTCGTGTGGTACTCCCAAGAGTTCCGTGAGAGGTTCGCTGACCCTTGGACGAGCT
TCACCGTGTCGTCCGGCATCACGATCCGGTAAATCTTCGAGTGGACGGTCTTCCGGTTCTTCAATCGAATCG
TGAGTCGATCGTCTTGTCGCAGTCGGACGAGCGACCTCGCCGTCGAGACTTCGTTCACCTGGTCGGCGTAGT
CTTCGGCGTTCCCGATGAGCACGTCGAGTGAGCCAATGTCGTACTCGGTCAGCATCTTCACCATCAGGTCGG
GCGTCTCCGCGTACGTCACGGCATCCACGTGACGAGCGCCTGCGAACAGATCAAGAAAGTCACTCCGCTCTT
TCACCATCGCCAGGCGGTATTCCGCCGCACCGGATCCGTAGAATTCCGGCATATCCGCAAACCTGTCAAACG
AGATATTCGCCGTCAGTTCGTCCATACACACGGTAGCAGTCTGAAACGAATAAAACCGTGCCGTGGAGTGAT
AGTGGAGCATTCGTCATCGTAACTTCGACGCCGACGCACGCCTCATCGGGCTGGAAGCCGGCCAGTGATTGA
GGAATGGCGCGAAGCTTCCGATGGACGGCAGCGTAAGCCGACGCGATCTGAAGGGGCTGATTTCGGTGAGGG
TGATTCGACAGGGGTGCAGGACCTGAACTGATCGTGTTGTATGTGATCAATTATCTTTACCCTCAAAATATG
CTGGTAGTGGGGCTAATACATCTGATATTGCGGACACCCCCCGTGTCTCAGTACAAGTCGATACGGAATCGT
CGACTGCGAAAGCACCCAGCGAAAACTCCGCTGCGGATCCATCGACTAGGAGACGAGTATCGAGTGAGTCTG
CATCGCGCTGTTGTCGATCTCTCTTAAATCATAACCAGGT

AAATTCAGGAGCCGGAAGGAATTCTTGGTGAGGTAAGATTCTGTCCCGCGCACAGGTTGCTAGAGCCGGTCA
GGTAGTGACCCTGTTGCGCGGAGCTAAGGGCTCGCAGGCATTTGAACTGTGAGTATAATGCCGCCTCCCAGC
ATTGAACCAGGGGTTCAAATGCCTCTGGTTGTTTCATTTAGTTATGACAAGCTTGCCGCCTAAGGTCGAAGC

CCTCCACAACCGGATTAAAAAATCAGAAGTACTGCCACAAGATGACAAAGATGCGCTGTTGCAGTTCTCGGA
CGAACTCGGTGCTCATAACTATTCAACGGGCAGGCGGGTGAAACTCCTACAGCACTGCACGATGATGGCAGG
GGATTCGGAGAAATACAGCCCGGATCAACTTCCCGAGCCAGATCTCGTCGATATGATCGGTGACACGGAGAC
GGAGAAAAAGAAGGCGAAACGGTACGTCAGCTGGATCAACGGAAACTACGATAGCGAGGAATCAAAACGGGA
TCACCGAGTCGCACTCCGGATGTTCGGGGGGCACATTACTCGCGGGGATCCTGAAGACGAGAAACCATACAG
TGTCGAATGGATCTCGGCCGATCTCCCGGATGACTATGATCCAATTCCAGACAAGACGAAAATGTGGTGGTG
GGATGAACACATTCTCCCGGTCCTGAATAACGCGAAGTATGCCAGAAACAAGGCGGCGGTCGCAGTTGATTG
GGACTCTGGAACTCGCTCGGGTGAATTCCGGTCGATGAAAGTTGGTGACGTCGGGGACCACAAATACGGGAA
AGAAATCACAGTCAACGGTCGGCAGGGGCAACGGTCAGTCACCCTTATCACATCCGTTCCGTACCTCCAACG
CTGGTTAGAGGTTCACCCAAAAGGAGATGATCCAGAAGCGCCGCTCTGGTGCGACCTGGATACCGGGACGCG
AGGTGAGTTACAAGATGAAGCAGAAAAATGCTCCGAAAGCCTGTTGAGCGAGCAGTAGACATCGGGGAGGCC
CCGCCACACACCACCCCCCC

DL31

Replicon: DL31_Contig114 Shared region: 59595..76202

GTGTTTTCCCTGCTCCGGCCGCCATGGCGGCCGCGGGAATTCGATTGGCTGGGCTGGAACGAGACCACCGTC
CCTCGACGGATCGTTCGCGACGAGATCATCACCCAGCTCACGCGGTCGCCCTCGACGACGGCGGAACGATCA
GGTGAGGGAGCAACCAGCGAGCCAGACACCTTCGTCGTGAAGCCAGTTCGAGAGCTCCGAACGCCGTAAGGG
CTAGACTTAACCAACACCTCGGACCCGTACCGACACAGTGTTGGTTAACGTTGAAGGTGGGTACACTCCCTC
GTCCCCGCCCGCCGCCCTGATACTCGAGTGAGTGAATACTGGCGGCCGGTCGACTCGAATGGCCGTCGCGAC
GGGTTTTATCGGGCCGCTGTTCGTCGAGAGCGAGTATGGTCGTCGTTGAGGAGGTCACCAAACTAAGCGAGC
GCCGCACCCAAACCTCGTCCGAGACATTTCCGGATGAGTCGCTGGCGTCGATTCAGACGGCCGTCGAGGCCG
TCCCAGCAAGCGTTTTCGACGGATCGACGAAACACACGGAGGTCGGCGGGTTCGATGCCGCGATCGCCGACG
ACCTCTCGCAGTACGAGTACGAAGGGGACGCCGCTGGTGGCTTCAACCCGAACTGTAAGCTCTGGTCGCATC
AGGGGTTCAACTACAGCGTCGACCTCTACGACGCCGACGCCCGCATCGCCATCGAGGTCGAGAAGAGCGAGC
GAAAGAACGTCAGCGACGACCTCCTCAAGTTCCAGAAGGGATACCGCACCCAGAAGGACAGCCGCCCGAAAA
TCGAATTACGCTGTCTGGTGGTGCCGGTGAACTATCTTGGCCGGCATAACCTCTACCAGCACAGTCTGACGA
AGCTCGACTTCATGAAGGGCGTCCTGTTCATCGACGACGTCGCCGTCATCGGCTATCGCGACCCGCGACCGG
ACTGACGCTTCTCGCGCAGACTGTTTGTCGGCGCCGGAGGCGTGCAGCGCGCGACAGCGTGCGCGGCGTGAC
CCATTATGGCTGGTCCAGATCCGCACGACGACACGAGTAGCGACCACGACAGTTCGAGAGGACAGCTCGCAC
AGGACCGCGACGCCGCCAGGCGTCGACACGGCGGA

CCGGCCGCCATGGCGGCCGCGGGAATTCGATTGAGGAGGCTATAAATGGTAGAATCGGGATCAAGGCGGCGC
TCGACGCGATCCTTTATCGTGGCGACCGAATCGGACCCCTCTGCCGAGTCATTAAAAATCTGGTATGTGTTT
ACCCCGAGGACTCGACGGATGGCACAGCGTGAGTCCGTGAAATCATCTCGTTGTCTCGGACAATGCGCTTGA
GCTTCTCGTAGGGATTGCGTCCCTGCTGGCGCGATGTCGCCAGCAGGGACAGCAACGTCTCGTGAACGAACA
TCCCGCGGTCGTTGCGGAGCGTCCCGATGATTTTCCGGAGAACAACCGGCTCACGAAGCGCATTCTCCGCAG
CGTTGTTCGTTGGCGAGACTGCTGGCTCACCGATGAAGGTGAGCCAGTGGTCGATCCCTCCTTTGATCTTCC
CGAGCAGTGTTGCCACTGGGTCGTCGGTAGCTGAGCACCTAACGAGCGATTTGAGTCCGTTCTGGCTTGATC
GGTGCATCTGTGCTCGCTCACGAAGCTCCGGGGTCGGTCTCCAGCCACGACTGGAGACCGACGTACATTTGC
GTGAGATGCCGGTGAATTGGCTCTCCGTCCTCGTACTTGTCAGCGACATCTTCCGCTTCGCGGAGAATATGT
GCCCAGCACCGCTGGAGGTTGCTGGTGAATGCCGGATATGCCGTCCAGCCATCGCAGACGGCCGTTCCCGCG
AAGTCCTCGCCGAGGACTTCTGCCGGAACATCGCTTCCGCGACTCTCTCTGACCGCGTACAGCGTGTGCTCG
TCCGTGGTGAACGTCCACATCCACGCCTGTTCGCCGTCGCGTTTGATTCCCGTCTCGTCGATGTGAACAACG
TCAGCGTGCTGAATCCGTCGGCGGATCTGTTCGTATTCACAGCGACCGGCGCGCGCAGCGCGCTCGGTCGCG
TGCCACGCAGATGCACCTGAGAGTTCGAGGCCATGCAATTGCTCGAAGCGGTCGGCGATCTTCCGGTAGGGG
AGCGGTGATCGTATCTGGAAGAGCGGCTTGGGCGATGACGTTCACCCCGAACTGCCCCTCGTCCGGGCAGGT
CCGGGGTTGTGAAGCGGACAGTCCTCGCCTCCCACCAGAGATA

AAAATCAATTCTTCACGCGTTGGGAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGCGAATTCACTAGTGA
TTGGACTACGGTGGCAATCTCTATCTAATGGGCGCGCGTCACCTGTGGGTCGATGATCCGAATGTACAGGAT
AGCTGCTCTACCCAAGATCCAACAGGTGAAAACGCCTATCAAAGTGACGACTACTACGGACACGTAAAACAC
CACTTCCAGGACTACGACGCCGCTCTCACAAATATTGAGGCTGATGATCAGGAAGGACAACCGCAGCGCGAT

GGTTTTACTGACACGATTGTCGACGAAAACGGCTATATCGAGGGATATGTGGATAGTAACGGTATTGACGAT
ATGATGGCGAATAATCTCGAAGTCCGTAAGCGCGGCATCACTACCGGTCCAACCACGGGCGTAATCGAAGAA
TACTTGGACAATTACTGCAGCACGAGCGTCACGAGACGCAGTCTTCTGCATGTATCCAACGAGCAACAATCC
GGAGATTCTGGTGGGCCAGTGTACGACCGTGACTACTTTGAAGGAAATTACTACCTCTTCATGGTTTCTCTC
GCCACCCAAGCAACGGGAGCCTCAGAGGCTGTCGGCTCAACAGCCCATTCAATGGCAAACAACCTTGGGATT
CAGTGGAGGACTCGGTAGTTAATCCGGAAGGGTGGTTAAGTACCACCCTCCAAGAGTGAGTAATATGACAAA
GATCCCGCCAAGCAGGAATGCGATCTCGGGGAGTGCGAGAAACTGAACAAGATTTTTCATAACCAAGACAGG
ATTTGAGGGATCCGTATCCGTATCAACGATTGTTTCTGGCCTAAGGTTACTAAGTACCCAACCAAAGAGGAA
ACTCGCAATACCAAGGATGCCAGTGATTCCAAGTGATTGTTTGAGAATCTTGCGCAGAGACCGCTTCTCGTT
CGAATTGCCGGCAAATAATATAATCGAACTACTAGTCGTGCGTACACATTCATTTCACGCCCTCTATCTGAA
TACCAGATATCAATAATTTCTATGAGTCCAGCCTCCTCAAGCTTATCAAGGTGATAACTTTGCGTTTTGGAC
CGACATTATCGAAGTGATTCTGCGGATTATCGCTTCGCCGGTCTCGGCGGTTCCTCCT

Replicon: DL31_Contig114 Shared region: 111139..146012

CCGGGGTCTTTCGTGTGATGGCTCATCTTGACTATTGGTGGGTCGTGCATCTCCCCGAGCCTGTCATAATCA
AATGTATTCCTTCAAGCTTACATTCGAATTGCAACTCCACTGTTTTTTCACACACAAGGCGGTTTGTCGCGT
GGTACGATAGCGGGCCGTTCGATATCGGACGGATGACGATGCGGTCGCTCAGCCGTCTTAAACACCGTAACA
CGTGGGACAACGCAGGCCATCACGTGTGGGAGCAGAGCCGGGAAGGCCAGAACGCAGGCAACGGGAGCGCGA
TGCGGTGTCCACCGCTCGCCATCCCGTACGCGACGGACTGGGACCGACTCGTCGAGGTAAGTCGGCAGTCCT
CACAGATCACGCACGCTGACCCACGATGTACTGATGGCTGTGCGATCCTGAATCTCACAGTGGCTGGCCTCC
TTGAATATTTGGACACGCCGTTGCAGGACGCTCTTGACTACGTCGGTGCGGACCCGCCTGAGGAGCTCGTGA
TGTCACTCCAACCGCTCGCTCGTGGCGACGCACCTGGAACGCTGGAAACGTCAGGGTACGTCGTGCATACAC
TGCATACGGCACTCCACGATGGCCTCATCGCGGACAGTGCCGAGGACGCTATTGTGACGGCTGGTGAATCGT
GGCGGTGATACGGATACGATTGGCGCGATTACTTGGCGCAGTCGCTGGCGCGCGGGTTCGGAGCGTCACATC
TTCCCGGATCGATGGTTGGGTGCTATCGTCGATGTC

Replicon: DL31_Contig114 Shared region: 664513..683577

TAAACAATTCATCCACGCGTTGGGAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGCGAATTCACTAGTGA
TTGTATGATGGATCGTTGACCTCGGTAAACTGGTGATCTGGGACAACGTTGAGTGTCGAATCTATATACTCG
TCTTCGAGGCCGTACATACGAAGATGGCTCTGTTGAAATCCTCAGAGATGTACATGTTCGCCGTGTAATTCT
ATGGGATGAAGTCCCTCCCAAAGTCGCAGATTCTCCGGTTTACTGAGAAGGCGATCCATCTGGCACGCCGGG
CGGTCTCTCGGTACTCCTCGAAATTCTCTAAACACCGCTATACACTTCCGCAGCACGTTGTTCTGCTCTGTC
TCAAAGTTCGGAAGAACACGACCTACCGTGGTCTGCTTGACGAACTGATCGAGATGCCACGCATCCGTCGTG
TTCTCGGGCTAGCCGAACTTCCTACTGCTTCAACGCTTTGTAAGTCGTTCAGCCGGCTTGATATGGCTGTAT
GGCGTGTCATATTGACTCTCTCAGCGACACTACTTCCGACAAGCGGCGTTGTTGGTGTTGATGCGTCAGGGT
TCGACCGCAGTCACGCTTCGAAACACTACACGAAACGCGCTGAACTCACGATTCAGCAGCTCAAGGTGACGT
TACTGGTCGATGCGAAGGTAAACGCGATACTCGATCTACACGTAACTACGACGCGGAAACACGATAGCCAGA
TCGCTCCGTCGTTGATCAAGCGCAATCCCGACGATATTGACGTTTTGCTCGGTGACAAAGGGTACGACGATT
CAGAAGATCAGGCGGCTCGCCCCGGCAACACGAAATTCGACCACTGATCAAGCATCGTGAGTTCACGTCACT
CCATAAAGGCTTGGAACGTACGGCTTAGACCTGATCCTCTACCGTCAGCGGAGTCCATCCGAAAACTGGTCA
ACTCAACACTCAAGCGGAAAGTACGGGGTGTTTTGTCCGGTTCAGGGCGCTGGTGGAACCGTTTCCTGGAAC
CAACTCTGA

TTAATTTTTCCACGCGTTGGGAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGCGAATTCACTAGTGATTC
GATACTAATGTCCTCACTCAACTGGGCTGTTGACAGCTCATCTTTGTCGATCGACGGTGAGAGCCGATGAGC
AATGAGAGCAGTCGGGTTCAAAGTGGGCTGACAAAGCAAGTTGATGATGTCCTTACTGCAGATACTGACTGG
ATCACACTTGCGAACGAACTGGACGTGAGCCGCTATACGCTACGGGACGCACACCCAGAGTGGAGTTCGTCG
CTTCCGTTTCGGCCAATGTTTCTGGCGTATCTGTGGGCAACTGTCGAGCGTGAATCTCTGTCAGGAATCCCA
GAACGCCTCTCTGACCGACCGGAACTCGCCCGTGCATTTGGGTTTGAGATGGATGATCTCCCCTCAGGAAGT
AGCTGTAAACCAGTCCGGCTTGAAAGCCGATTCAGAAAGTTACAGACGGTCGTCGAATCAGGTGCTGAAGAG
ATCCGCCTGCTCGCGGCTGAACGAGGCGCACCAATCGGGAATGATCTTCTCAAAACAGCGGACGACGAAGAC
AAACAGTCGCTGTCAAATCGAACCGTCCAACGCTTGCTACGGAAGAAGGGGCATCAGGTGCTTGATGAGTTG
AAGTCGGTAGCCATCCCTTCAATCTCACTCTCTCGCCCGGATGACGCGATCTACGACGACGATGAGTTACTC

GTCTTAGAAGCAATCGCGTCGATCAAACAGAAGGCAGCACACGATTCGGGCCAGAAGCTGGGTGACATGAAA
AATCCAGACCCAGATATTGATGACCCGTTCTACGAGGACGGCCCATCTGGTGAGACGCTGTTGGAAGCCCTC
AAGCAGATGTCTATCGAGGAGATTGCGACTGTACTGAATTTCGCTCTCCGGAAAACCTACACACGCGCGAAA
CCCCGAATCAGGAGCTCGAACACGGGAACGGCTCACGGTTTGGGACTCGTGCGAAAGTCGCTCTGGATATGA
CGTACGTTGCCTACTATGGCGATCGCGACGAGATGGAATGGTACAGGGCGCCACCTGACGAAAAGAGTACAG
TTGGTGTCACAAGTTTGCGACGGTCGTGATCGTCGGCGAGAACACCCACTA

TAACATTATCACGCGTTGGGAGCTCTCCCATATGGTCGACCTGCAGGCGGCCGCGAATTCACTAGTGATTGG
TAGTGCTACGCTAAAACAGTGCCTCAACAGTCGGCAACGAGATGTTCCTCGAGGTCGCGCGTCCAGTACGTC
GCGAGGCCACCAGGGCTACACCGCTCGCACAGCTGCAATGGGCCTTCGAGATGACCGAGTTCGACGCGGAAG
CGCGTCAGCGCCGCGAGGGCGTCGACGACGAGCCCACAGCCGTCGCAGGCGTAGTGATTGCGTTCGTCGGGA
TCCAGTTCCGTCTGGATCGCACAGTCGACACAGATGGGGTGAGTGACCCGCTCGTCTTCGCCCCAGGTATGC
GCCTCGCCACTCTCGGTACCGGGCGGGGCGTCACAGAACGCACACCCTGTGAGCGTCTGCTGCATCACTCGC
CCTCCACGTCGGCGTCGCGACCGGCTGTCCAGCCACGGTGGAAGACGGCCAACTGCCTCGCCGAGTCGAAGC
TAGTGCCACTCGGCTCGTGAACGAGGACGCGATCCTCGGGAACGGGTGTGACGACGTCCGCGTCGATGAGTT
CGCTGACCAAGTTCTCGACCGTTGCGTCGTCGACGTCCGCCGTGTCGACGCTGGCGGCGTCGCGGTCCTGTG
CGAGCTGTTCCTTCTCGAACTGTTCGTGGTCGCTACTCGTGTCGTCGTGCGGATCTGGACCAGCCATAATGG
GTCACGCCGCGCACGCTGTCGCGCGCTGCACGCCTCCCGGCGCCGACAAACAGTCTGCGCGAGAAGCGTCAG
TCCGGTCGCGGGTCGCGATAGCCGATGACGGCGACGTCGTCGATGAACAGGACGCCCTTCATGAAGTCGAGC
TTCGTCAGACTGTGCTGGTAGAGGTTATGCCGGCCAAGATAGTTCACCGGCACCACCAGACAGCCTAATTCG
ATTTTCGGGCGGCTGTCCTTCTGGGTGCGGTATCCCTTCTGGAACTTGAGGAGGTCGTCGCTGACGTTCTTT
CGCTCGCTCTTCTCGACTCGATGGCGATGCGGGCGTCGCGTCGTAGAAGGTCGACGCTGTAGTTGAACCCTG
ATGGCGACAGGAGCTTACAGTTCGGGTTTGAAGCCACCAGCGGGCGTTTCT

tADL

Replicon: True-ADL Shared region: 1940216..1952775

CGCCTCGATGTCTGCGAAGAGCGTACGTACTCGTCTGCGTGTTGGTGAGAATGTGCGGGCGATGCGACTGGA
ACCGATTCCGCACTGTGCCATAGGTACATTCCCGTTCGTCGTTGCTCGCGGGCGGTTCGGTGACCCCGCACC
CCTCTCAGGGGTTCAACGAACAGGACGGCGTACCGTTCGGCAACGTATTTGACAGTTGCAACGTACTGTCCA
TATATGTCGGGATGGCTGTACTGAACCTCTCGAACATCGGCAGTTCGAGGACGTGGTAGAGGACGTCTTCCG
CAACCTCGGCTACGAGAACGTCCGCCAGGCCGACCGCACGGCTGACGAGGGTCGCGACGTCATCATGGAGGA
GGTCGTCGACGGCACGCGGCGTGCGATCATCGTCGAGTGCAAGCACACGGGGACTGTCGGGCGGCCGGTCGT
CCAGAAGCTCCACTCGGCGATCGCGACGTTCGACTTCGACGGTCCCAAACGCGGAATGGTCGTCACGACCGG
CCGGTTTACGAACCCTGCTCAGGAGTACGCAAACCGCCTCCAGCAAAACGACGACA

CGGGTTTGAGGATGGAGCCAGCACGTCGAGAAAACGGTCGACATGAACTCGGACTTCAAAAAATCGATAATT
GTCTGCAGCTGCGTTGCTCGGACCGCAACCTACTCCCAGGACTCCGGATGACTCTTCTAGGCGGGTGAATAG
TAACTGCTGTCCTCGACGTGCTCGCTCTTTCTGTTCAAACCATTCTTTTGGAACCTGTTGAATTTGACGCCG
ACTCCCTTCGACCGATGAGTGACGATTCGAACACGATCCCTGCTAGGAGTGGCTCGAGCTTTCCCGATGGCC
GTGTTGAGCCCTGCTCCAGAGATGCCGATGGCTCCGAGTGATTTCATGAACTTTCGCCGGTCTATTCCCCTG
TTTGCTCGGACTGATCCGAGCCGAGATTGGTCTTCTGTCATGGGTTAACACCCACCAAATAGAAAGTATGGA
GACATTATTTCGCTAACTATTTATCCAAACCAAAACGATGAGTGGAGGCGAATTTGTTCATAAGCGCTACAA
TATTGCTAATACGCATCCGGTAATTATATCCAACCACTCTATTGGATAATACAAAATTCGATTCGAAGAGGG
GAGAGATTTATCCAGTTTCAATATAAACTTCCGGTTATGTCCGGTCCCCGAGTAGCGGAGCTGTTAAATCAC
CTTGCACACAGAGGAGATGTCCTGAAGAGTCTAGAAGAAGGACACTCTGACCAGCCGT

Replicon: True-ADL Shared region: 1234151..1243860

CGATCGTGGAAGGACCCAAATCCCAGCCAGACCAAATAGAGCACGTCGAGCGGAGCGACGTCGGTGTCTCAC
TCACTGTGCAACTCACTCGTGGGACGGGGACACGGGATCAAGACAAAATAACAGCGAAGGTGAAAGCGAAAA
CCTTGGAGGAGGCTCGAACCGACATGGAGACTCTTCGAGCATATATTGACGACCTTGCTGAGTCGACACGCC
AGATCCAACCAGCAGAGGTACCCAAGTAGCGAGTCGACCTGATTTTTTCATCGCCTCACCAAGGGTAGAGGC

GATTTAATTGAGTAATTCAGCAGATCAGTCCACAGAATCGAGTGGCCTCACCCCACAACAGCGTCTCGAATC
GTCGAACACTCGGCTCGTCGACGCAGGCATTGCGACGATCAAGGATATGGAGACGCTGCGGGCGTGTGTCGC
CTATGAAAATGCGAACCAACGCCGGGTTCTGATTCTCCATCGTCTCAAGCGACGAGCTGACGAGATCCGGGC
CGAAGTCGAATAGAAGAGCACTCACTGACTCATTCTGGGCCCACGCGATGTTAACCAACACTCAGCCGTAAC
GACGAATTTTGTTGGTTAAGACTGCTACTGATAACGATTTTTTTCTCCCCGAGAGGTGTGCGGGGAGCGAGT
TCTCGCAATGATTCAAAATGGCATTATCAGCACGTTCGACCAGTAGCAAAGCCAGCGAGATTGTAGCAGCTC
GAGAGACCGACTACGTCGCATTTCTCCACCGCGTTCCCTTCGCGTTAGATGCCTTCAACCTCGGTTTTCTCA
CCGGCTTTCGAGAAGACTGTACTTACCAGCAAAACCAATACACGGATTTGGAGTTGCCAGTTGGAATGCTTG
ACAACGACTTCCGGAATCCCGATCTCACTCGATACGTGGAGCGGTTTTTTCGGTACGAACCACAGGTCGGCG
GGGATCGGGTGATGCCTACAGCGTCGACGAGGTCGAGGAATACGTAGCTGCTGCTCGCGATATCCAAGCAGG
CTATCCAGGAATCAGAGTTGGTGATCGTCCCGGAGGTGCCGGTGCAGCCATTCATGCAGTACAGAGGACTCA
TCGTGGGAATACCTCACGGGGATACGCTGAACAGCTAGCTCACGAGTTTTCGGAGCGACTGAATGGCCGTGG
TCGACGTGTCATTCTCGAGGGGAGTCATTGACACTCAAGGGGATCAGCAGCTACTGTCATCGTGACGGGTGA
TCTCTCTGGCGAGAAATTGTGTCCGGGCCTGTGAACTTGGA

CTTTTTTGGGCAAGCCCTTCTTGAAAGCCGCTGAAGAATCGACCCCTGTACCGTCAGTCGTGATCGACGGTG
GCCGAATATCATTATCGAGCGCCAGCAGTAGGTCGCGAGCGTACTTCTGTCGACACTCAGTAATTGCTTTCG
AGCGTTCCCCTCGGTCAGTTAGCTCATCCTGGACGAGCCCTGCGAGTGTCCGTTGGTCGAACGTCGGAAACA
CCGGCCAACGCTCTGTCGGCGGGTCCATCAGCTGGCGATAGCTCCGCAGCGGCGTAATAACCGGGTCGGGGA
GACTCGCGGCGTCCCACTGCTGTTTCTTCCGGTAGACGTCCATACTCCCGTCGTCGAGCGAAAGTTCCTCCC
AGCGGACGCCACGTCGACGCGGATCGTTCGGGTCCCGAAGAAGTTCCCCGACCCGAACGGCGGTGTATGCTA
GGACGAACACCAAGGCTCGGTCTCGGGCAGCTCTCAGGGCAGTATATCTAGCTTGTTGCTTCTCACGTGGGT
CGACATCCTCTGAGAGTGTAGTGAGTGCCTCGATGGCATTACGGGCTTGTTCGTCGACGTGACGGGTAAGTG
CGTGACGCTGCTTCGGAGGTCCAAGCCTGCTGGTCACCGGGTCTTGCGACCGTCGTCATCAGGGAGTGGCGC
GGTCGCACTGGCCCGCTGGGCGTAGTGTGCTTCAAGATAGCCTTCGTTTGACACACCAGCCGCACCACGTCA
AGTTTACGAACGACACAAAACGTTTCACGGCCTCTCTTCACCCGGTATATCCTGTATCAAATTTTTCAGCTG
ATGTTACACTGTCTAAAACCTATTCGATGTGTTTGATTGTTTTACTAAGGATAACTCCCATAGAACTTATTT
TCTAGGTTTTGTTTTTGACACCGCTGATCCCTACGACTGCGTATAGAAA

Replicon: True-ADL Shared region: 1243876..1250380

CTTTTTTGGGCAAGCCCTTCTTGAAAGCCGCTGAAGAATCGACCCCTGTACCGTCAGTCGTGATCGACGGTG
GCCGAATATCATTATCGAGCGCCAGCAGTAGGTCGCGAGCGTACTTCTGTCGACACTCAGTAATTGCTTTCG
AGCGTTCCCCTCGGTCAGTTAGCTCATCCTGGACGAGCCCTGCGAGTGTCCGTTGGTCGAACGTCGGAAACA
CCGGCCAACGCTCTGTCGGCGGGTCCATCAGCTGGCGATAGCTCCGCAGCGGCGTAATAACCGGGTCGGGGA
GACTCGCGGCGTCCCACTGCTGTTTCTTCCGGTAGACGTCCATACTCCCGTCGTCGAGCGAAAGTTCCTCCC
AGCGGACGCCACGTCGACGCGGATCGTTCGGGTCCCGAAGAAGTTCCCCGACCCGAACGGCGGTGTATGCTA
GGACGAACACCAAGGCTCGGTCTCGGGCAGCTCTCAGGGCAGTATATCTAGCTTGTTGCTTCTCACGTGGGT
CGACATCCTCTGAGAGTGTAGTGAGTGCCTCGATGGCATTACGGGCTTGTTCGTCGACGTGACGGGTAAGTG
CGTGACGCTGCTTCGGAGGTCCAAGCCTGCTGGTCACCGGGTCTTGCGACCGTCGTCATCAGGGAGTGGCGC
GGTCGCACTGGCCCGCTGGGCGTAGTGTGCTTCAAGATAGCCTTCGTTTGACACACCAGCCGCACCACGTCA
AGTTTACGAACGACACAAAACGTTTCACGGCCTCTCTTCACCCGGTATATCCTGTATCAAATTTTTCAGCTG
ATGTTACACTGTCTAAAACCTATTCGATGTGTTTGATTGTTTTACTAAGGATAACTCCCATAGAACTTATTT
TCTAGGTTTTGTTTTTGACACCGCTGATCCCTACGACTGCGTATAGAAA

GGTTCGAGATCCTTCCTTCCAGAACTCTCCGAAAGATCGGTCGTGGTGTCGACGAAAGGCCGCGAGAAATGA
CCGCACACCGTGATCACTGAACCACTGGTGGGTTGGTTTCTCCGCCGTAGGATCAAGGCCTTCTGCTTCGAG
ACACGGAGCAACCTGCTCCCAGTAGAGCTGTGTGAAGTCTTCGAGTGAACAGGCCGACCACCGGACATCGTC
GAACGAAGGCTGCTGTGGTTCGAGCTCGCTTTCCTGATCGGACTCAGGGTGACTCATTGGGTTATGCCCCGG
CTGTTGTCTGGATGAGGTTCTGAGAACAGATTTCGAAGTAGAAGCCTGCTCGCTGCCTCGTGTGGTAATTCT
AAGAGTCCCATATTCTGGTTGTTCAGTCATTCAGTAGGATTAATTATAAATTTATGGGTTACGCTGACTACG
AGAGTCTGCTAAACCCATGGTATATTCGGCGCGAAACCACGCCACATCAATAAATCACATATTGCGATATAG
ATTTCTGGGCAGTCACGCTGCGAATGGCGGCTCTCACGTCACGAGGTAGACGTTTTCTCGCCGGTATTCGCG
GCAGATTAAGGGTAAATCAGTGTAAAAATCCGTCTCTTAGAACTATCTCGTTAAATACAAGCAACTAATTGC
TGAAAACTAAACAACCGCGAAAAGATGTCTCTACCGGCTAGCAGACCTGGAACATATAATATAAACCAACAG
AGTGTGTGGTTGTACGGCCTGTGTTGGTTAAGTCATATCAGCAATCCCGATTTGAAGATGAGCAACTAGATG

TCTAAAAGGCTGATTTCTTTCCAGACAGCTGAACTCATGTATCACAAAATGTTCGTTTAAATTATAGTGCTC
TTAATCGAGATTTATATCAGGCATGTCGTACGATAATTTAGACAGGTCAGTTAATCAACGCACTACTCGATG
ACGGGTCGGGCCAGCCCTTCGAAGCCTTGGAGAGGGGCTTGATGTATCAGTGACCACCGTCTTCAAATCACA
TCAACGATTCTTCGAAGCAGAAGGTGTTATTGAAAGGCTATACACCGGTTCATTAACTACGGCGAACTGGGC
TACGACGTACCGCATTCATGCAGCTGATGTCGAAGGATCCGCGCCTTGCAAAATTGTCGACAGCTTTCGCGA
CAGGTACGATATCCACTGTTCTATTGAAGGTCACCGTGATTAGACCTTATTGGCATTGAAGTCCGAGATACA
CGGAATGAACCGCTCGGAATCCAAGAGACTGCTCATCTTGTGCCGTGTGACCCATAGTAC

Materials and Methods

Sample collection and processing for metagenomics from Deep Lake. Water samples were collected from DL (68°33'36.8S, 78°11'48.7E), Vestfold Hills, Antarctica between November 30 and December 5, 2008 (Fig. S1). Water was collected by dinghy from above the deepest point in the lake by pumping water directly from 5, 13, 24 and 36 m depths into 25L drums and immediately processing samples on-shore by sequential size fractionation through a 20 µm prefilter directly onto filters 3.0, 0.8 and 0.1 µm pore sized, 293 mm polyethersulfone membrane filters (12,13,17-22). A volume of 50 L was filtered for 5, 13 and 24 m depths, and 25 L for 36 m depth. Samples were preserved in buffer and cryogenically stored, and DNA extracted as previously described (12,13,18-22).

Isolation, growth and genomic DNA extraction of DL haloarchaea. tADL (NCBI taxon ID 758602), DL31 (NCBI taxon ID 756883) and DL1 (NCBI taxon ID 751944) were isolated from DL surface water collected December 2006 (tADL) and November 2008 (DL31, DL1) (Fig. S1). Pure cultures were recovered from water samples using an extinction dilution method (23) and DBCM2 medium (24,25). All cultures were incubated at 30°C. Repeated rounds of limiting dilution titrations produced pure cultures, as assessed by microscopy and 16S rRNA gene sequencing (16). For large-scale cultivation, cells were inoculated into 200 ml of DBCM2 medium in 500 ml capacity, cotton-wool stoppered flasks. Cultures were shaken (100 rpm) at 30°C until late exponential phase and cells harvested by centrifugation (5,000 rpm = 4066 x g, 15 min, 4°C, Sorvall GSA rotor). The cell pellet was resuspended gently in 2 ml of a solution containing 20% (v/v) glycerol and 2 M NaCl. Cell lysis and DNA purification was performed using Qiagen genomic tips (500/G), and the manufacturer's protocol for the extraction of DNA from bacteria (Qiagen genome DNA handbook). The resulting DNA was checked for quantity and quality by spectroscopy ($A_{260}/A_{280} \geq 1.95$) and by agarose gel electrophoresis. PCR and sequencing of the 16S rRNA genes was used to confirm identity and purity of the DNA preparations.

DNA sequencing. Metagenome libraries for pyrosequencing were constructed using DNA from 0.1 µm filters (at 5 m, 13 m and 24 m depth), 0.8 um and 3.0 um filters (at 24 m depth) or pooled DNA from all three filter sizes (at 36 m depth) using the RAPID protocol (Roche) and sequenced using 454 technology (26) on a 454-FLX machine using Titanium chemistry. Illumina sequencing (27) was performed from libraries made using DNA from 0.8 um and 3.0 µm filters from the 24 m sample only, using recommended protocols (Illumina), and were sequenced on the Illumina GAIIx using paired-end 76 cycle reads. Pyrosequencing of PCR amplified V8 region of small subunit (SSU) rRNA genes was used to generate microbial community profiles. The 454 adaptor-added SSU rRNA gene primer set, 926Fw (5'-3'AAACTYAAAKGAATTGRCGG) and 1392R (5'-3'ACGGGCGGTGTGTRC) was used in PCR to amplify the V6-V8 region, with 5-bp barcodes incorporated into the reverse primer to multiplex samples. PCR amplicons were sequenced by the DOE Joint Genome Institute, using the Roche 454 GS Titanium technology as previously described (28). Sequences were analyzed through the Pyrotagger computational pipeline (http://pyrotagger.jgi-psf.org) for quality trimming, clustering to operational taxonomic units (OTUs) based on 97% sequence identity, and

taxonomic assignment by blastn against the Greengenes database (29). Singletons and potential chimeras were removed to minimize PCR artifacts. Draft genomes of the DL haloarchaea were generated at the DOE Joint Genome Institute (JGI) using a combination of Illumina (27) and 454 technologies (26). Briefly, for each genome we constructed and sequenced an Illumina GAii shotgun library, a 454 Titanium standard library and a paired end 454 library with an average insert size of 8-10 kb. All general aspects of library construction and sequencing performed at the JGI can be found at http://www.jgi.doe.gov/. The initial draft assemblies were generated using a Newbler/VELVET (30) hybrid approach. Manual finishing of all genomes was then performed at Los Alamos National Laboratory using a combination of computational tools, as well as PCR fragment subcloning and PCR-based primer walks.

SSU rRNA gene profiling. The taxonomic affiliations for SSU rRNA gene V6 tag sequences contributing to >0.2% of all sequences derived from 5 m, 13 m, 24 m and 36 m depths combined in DL were resolved by alignment of tags and their five nearest blast hits from NCBI to the Greengenes 2011 alignment and insertion by parsimony into the guide tree. Class level taxonomic affiliations of SSU rRNA gene V6 tag sequences comprising <0.2% of all sequences derived from 5 m, 13 m , 24 m and 36 m depths in DL were derived from blastn against the Greengenes database. Phylogenetic tree construction was performed using the software package ARB (31). Full–length SSU rRNA gene sequences from the four DL haloarchaeal genomes were aligned to the 2011 Greengenes reference phylogeny (32) and incorporated by parsimony into the reference tree consisting of 408,315 sequences. Reference outgroup sequences from major *Archaea* groups, along with close neighbours of the four DL haloarchaea were chosen for *de novo* tree construction using the Neighbour-Joining algorithm with Jukes-Cantor correction.

Fragment recruitment, highly degenerate regions and reference mapping. Fragment recruitment (FR) was performed individually on each sample and carried out against all publically available complete and draft bacterial, archaeal and viral genomes from NCBI, all organelle genomes from EMBL and the four DL haloarchaea (tADL, DL31, *Hl* and DL1) as references. The only eucaryal reference examined was a halophilic green alga *Dunaliella salina* CCAP 19/18 (http://genomesonline.org/cgi-bin/GOLD/bin/GOLDCards.cgi?goldstamp=Gi13840), where the raw draft sequence was masked for low complexity sequence using RepeatMasker version 3.3.0 (-noint). FR was performed independently for each reference using FR-Hit version 0.5.8 (33) (read coverage > 90%, alignment identity > 98%) and the resulting combined set of read assignments across all genomes was reduced to a non-redundant set by selection of the single longest and highest identity alignment per read. The genomic coordinates of highly degenerate regions (read depths > 3-fold median replicon read-depth) were determined by application of peak detection using custom scripts in R applied to replicon nearest-neighbour smoothed (radius 200 bp) read-depth traces. Sequences extracted from these regions were subsequently reduced to a non-redundant set by clustering with CD-Hit version 4.0 at 98% identity and annotated by BLASTX against Refseq_protein (e-value < 10^{-5}). FR was repeated on post-filtered DL metagenomic reads. Reference mapping was performed with GS Reference Mapper v2.6 for each genome against all DL samples

(98% minimum overlap identity, 100 bp minimum overlap length). Gephi version 0.8.2 was used to visualise and calculate statistics of networks (34).

SNPs and recombination. Paired-end Solexa runs from 36 m depth were aligned with Bowtie version 0.12.8 (35) to the four DL haloarchaeal genomes following recent methodology (1) with parameters: -t -n 1 -l 20 -e 80 --solexa1.3-quals --nomaqround --best --sam --chunkmbs 128 --minins 100 --maxins 500. Subsequent SNP identification was performed using Samtools version 1.1.16 with parameters: -c -C 50 -N 4. SNPs where mapped read-depth above 20-fold and variant frequency above 0.9 were considered fixed mutations within the DL population. PIIM v2.02 (36) was used to infer scaled rates of mutation and recombination ($\theta = 2N_e\mu, \rho = 2N_e c$) for each DL isolate replicon. PIIM's pre-generated likelihood lookup table limits read-depth to 50-fold, therefore requiring downsampling of our data. Downsampling by random selection was performed using Picard v1.77 (37) and BAM to ACE format conversion accomplished using Consed v23 (38). Rates were estimated in sliding 50 kb windows and resampling repeated 10 times with randomly selected seeds. Confidence intervals (95% CI) for θ, ρ were estimated by bootstrapping in R (N=10,000) using the boot (39) package. Rates estimation in PIIM was performed using the likelihood lookup table for $\theta = 0.1$, since as many as 25% of predictions resulted in $\rho = \infty$ when using $\theta = 0.01$.

CAI/CBI. Codon adaptation and codon bias indexes (CAI/CBI) (40) for each genome were determined by CodonW version 1.4.4 (41). Putative optimal codons sets were determined by a two-way Chi-squared contingency test of the two extremes of the principle trend of a correspondence analysis, where sampling of codon frequency was restricted to the primary replicon of each isolate.

Ortholog groups. Ortholog groups of the 17 haloarchaeal completed genomes (tADL; DL31; *Hl*; DL1; *Haladaptatus paucihalophilus* DX253; *Halalkalicoccus jeotgali* B3, DSM 18796; *Haloarcula marismortui* ATCC 43049; *Halobacterium salinarum* R1, DSM 671; *Halobacterium* sp. NRC-1; *Haloferax volcanii* DS2, ATCC 29605; *Halogeometricum borinquense* PR3, DSM 11551; *Halomicrobium mukohataei* arg-2, DSM 12286; *Haloquadratum walsbyi* HBSQ001, DSM 16790; *Halorhabdus utahensis* AX-2, DSM 12940; *Haloterrigena turkmenica* DSM 5511; *Natrialba magadii* ATCC 43099; *Natronomonas pharaonis* Gabara, DSM 2160) were determined using OrthoMCL version 2.0.2 (42) with parameters: -I 1.4. Granularity was adjusted for best agreement between the resultant clusters and known single copy genes (*e.g.* ribosomal proteins, tRNA synthetases) (4). The function of ortholog groups was inferred by homology to Archaeal COGs (43), where HMMER version 2.3.2 was used to create profile HMMs for each group from Clustalw2 multiple sequence alignments and ArCOG assignment decided by a simple voting heuristic. Core haloarchaeal genomic content was defined as ortholog groups shared as single copy genes between all 17 genomes and permitting one doublet (894 groups). DL flexible content was defined as those ortholog groups (68 groups) shared between at least one secondary replicon from each DL isolate and tADL.

Selection – K_a/K_s ratios (ω). Single copy ortholog pairs were identified between tADL and *Hl* (within population) and tADL and *H. volcanii* (between population). Clustalw2

was employed for multiple sequence alignment and ω was estimated by KaKs_calculator (44) using AICc model selection (-m MS).

Genome characteristics of niche adaptation. Genomes were interrogated for the presence or absence of genes that encode proteins and metabolic pathways that may provide ecophysiological distinctions between the four DL haloarchaea. COG scrambler (45) was used to identify statistical representation of COG categories and individual COGs between the four DL genomes.

Whole metagenome assembly. The whole metagenome of DL was assembled using Celera WGS (v6.1), with the recommended component pipeline for 454 Titanium data. An initial assembly was performed with default runtime parameters, with the exception of 3.0% utgErrorRate. Celera WGS estimation of genome size and consequently its effect on the discriminating statistic (a-stat) categorising early contigs as degenerate (repeats) performs poorly on datasets which do not represent a single organism. The tendency to over-estimate genome size leads to an increase in false positive rate of degenerates. To mitigate its effect, genome size was reduced manually by 2-fold and the assembly repeated post-overlap stage. This was repeated iteratively until the rate of change in degenerate assignment began to slow. The previous step was then taken as the final result. Genome size was reduced 4-fold from the predicted 88 Mb to 22 Mb. From 6,626,699 usable reads, assembly resulted in 16,551 contigs (mean length = 2736bp), 634 large contigs (≥ 10kb) totalling 12 Mb in total length and 15917 small contigs totalling 33 Mb in total length. An additional 61 large contigs (> 10kb) classified as degenerate (mean read-depth = 200) were also analysed. For assembled contigs the low complexity of the DL metagenome permits inference of replicon source by cluster analysis in a two-dimensional space composed of GC content and mean read-depth. Restricted to lengths greater than 15 kb, contigs belonging to each primary replicon of the three abundant isolate genomes were identified by stringent BLASTN assignment (coverage > 90%, e-value < 10^{-10}). Model based clustering using Mclust (46) was performed in R with background Poisson noise to compensate for the presence of outliers belonging to no cluster. As tADL contigs were clearly separated, clustering was limited to a subspace (20 < read-depth < 100). The unlabelled cluster (center: GC=0.63, RD=38.3) comprises 52 large contigs (> 15 kb) totalling 1.89 Mb in total extent and subsequently referred to as tADL-related 5th genome. Contigs attributed to the tADL-related 5th genome were mapped to tADL by CONTIGuator (47) using default parameters. Annotation was performed using SHAP (48) and predicted genes stringently assigned (e-value < 10^{-10}) to orthologous groups by Hmmpfam.

Similarity comparison. Average nucleotide identity (ANI) and tetranucleotide usage deviation (TUD) regression coefficients were determined between all replicons with JSpecies version 1.2.1 (49) along with the “tADL-related 5th genome”.

High identity regions (HIR). Long range HIR between DL haloarchaeal genomes were identified using NUCMER all-vs-all, where only regions with greater than 99% identity and longer than 5 kb were kept. Flanking regions (2000 bp upstream) were compared against Refseq_protein using BLASTX (e-value < 10^{-5}) where overall genomic content

was assessed by categorisation as insertion sequence, mobile element associated, or other. The 13 identified regions longer than 10 kb were reduced to a non-redundant set of 12 sequences using CD-hit (99% identity). To infer putative replicon source (lineage) for the 12 non-redundant regions, TUD regression coefficients were calculated between all 12 regions and all 9 DL replicons using JSpecies.

An all-vs-all analysis was carried out between 25 finished HA genomes (Table S9). ANI by BLAST was determined using JSpecies and the total extent in base-pairs of long shared HIR was determined using NUCMER. For HIR, only alignments greater than 99% identity and longer than 2000bp were considered, where the length criteria was chosen to minimise bias attributable to short, possibly well conserved mobile elements such as insertion sequences. The total length of the resulting alignments were then summed for each genome-pair. Two symmetric "comparison" matrices of genome-pair values for ANI and L_{HIR} were constructed. Due to the large value range and presence of zeros, L_{HIR} matrix elements were transformed by $x'_{ij} = log10(x_{ij} + 1)$ prior to further steps.

Fifteen metagenomes from hypersaline environments were obtained from the Sequence Read Archive (http://sra.dnanexus.com) in fastq format. Reads were converted to multi-fasta by in-house script and fragment recruitment performed independently for each long (>10kb) HIR using FR-Hit version 0.5.8 (33) (read coverage > 50%, alignment identity > 70%) against these metagenome readsets.

Matrix element definitions for each of the two matrices:

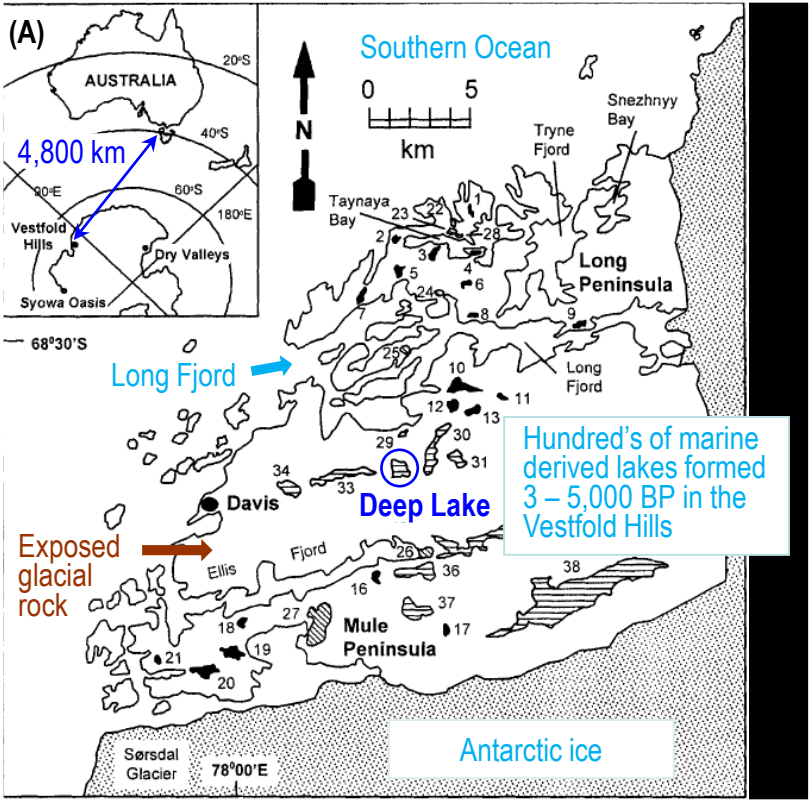
$$ANI_{i,j} = ANI(genome_i, genome_j)$$

$$L_{HIR}{}_{i,j} = log10\left(\sum_{i=1}^{n} length_{i,j,n} + 1\right)$$

Insertion Sequences (ISs). ISSaga (50) was used to analyse and manually annotate DL haloarchaeal genomes for ISs.

Lake viruses. Accessions for 55 completed archaeal virus genomes were sourced from the European Nucleotide Archive (ENA) (URL: http://www.ebi.ac.uk/genomes/archaealvirus.html) and retrieved from the ENA Sequence Version Archive (URL: http://www.ebi.ac.uk/cgi-bin/sva/sva.pl?&do_batch=1) as DNA fasta sequence. An archaeal virus database was generated from the 55 genomes, where hits from BLASTN with e-value < 10^{-5} were identified as significant. Fifteen contigs over 1 kb and 4 contigs over 10 kb were identified as possessing significant similarity to viruses associated with halophilic *Euryarchaeota* hosts (7).

PCR and DNA sequencing confirmation of HIR. PCR and Sanger sequencing was used to confirm that presence of HIR in each of the four DL haloarchaeal genomes, by amplifying and sequencing the boundary regions of randomly selected HIR from all four strains. DNA was extracted from cultures using the xanthogenate-sodium dodecyl sulfate extraction protocol (51). Primers were designed to ensure that the desired amplicons would include sequences immediately before and after the boundary of the HIR.


(A)
AUSTRALIA
20°S
40°S
4,800 km
90°E
60°S
180°E
Vestfold Hills
Dry Valleys
Syowa Oasis
68°30'S
Southern Ocean
0
5
km
N
Snezhnyy Bay
Tryne Fjord
Taynaya Bay
Long Peninsula
Long Fjord
Long Fjord
Hundred's of marine derived lakes formed 3 – 5,000 BP in the Vestfold Hills
Deep Lake
Davis
Exposed glacial rock
Ellis Fjord
Mule Peninsula
Sørsdal Glacier
78°00'E
Antarctic ice


(B)

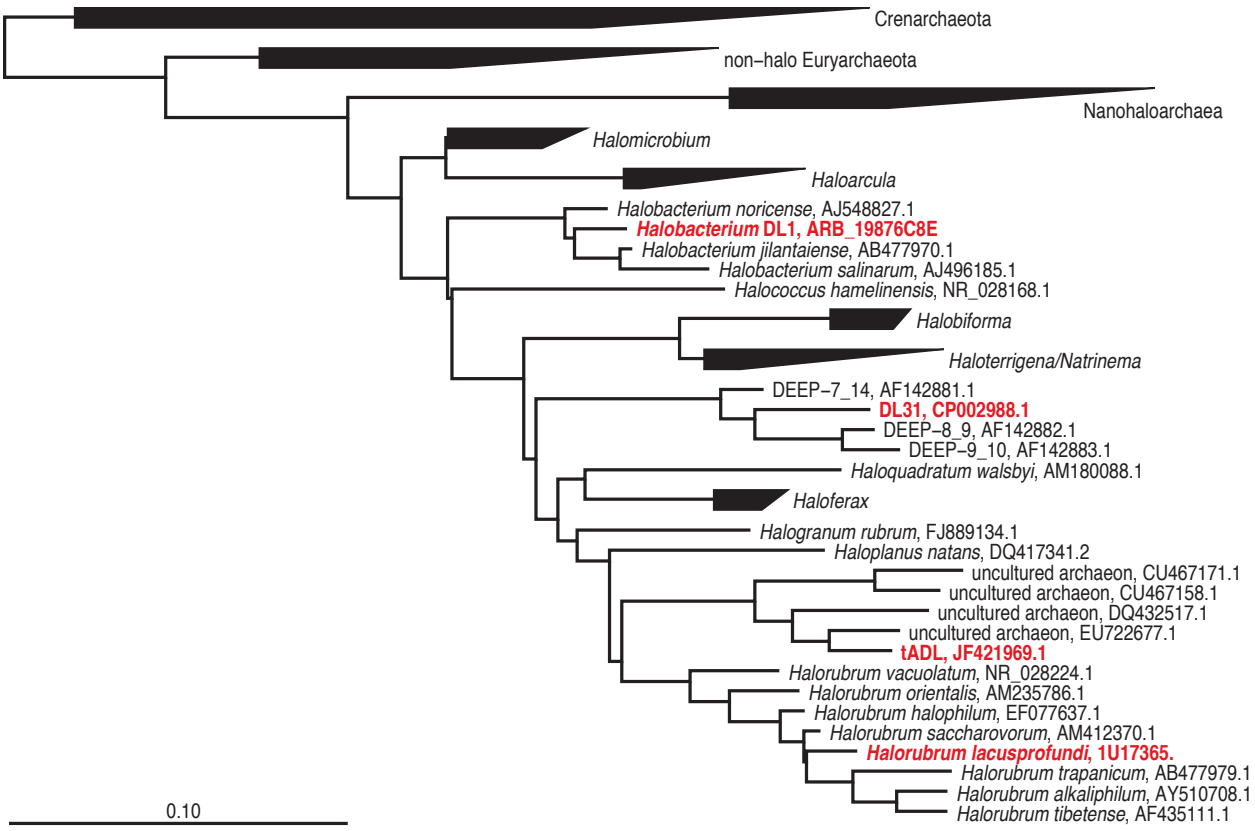

(C)


(D)


(E)


(F)

Fig. S1. Deep Lake expedition. **(A)** Schematic of the Vestfold Hills, on the eastern shore of Prydz Bay, East Antarctica, showing the location of Organic Lake, adapted from Gibson (11). The Vestfold Hills is approximately 400 km^2 in area and contains a remarkable diversity of more than 300 lakes which range in salinity from fresh to hypersaline (11,52). Most of the saline lakes were originally pockets of seawater, trapped less than 10 000 BP when the continental ice-sheet receded and the land rose above sea-level (11,53). Differing local conditions has led each lake to develop unique physical and chemical properties, and life in the lakes tends to be entirely microbial with low levels of diversity (52,54). Deep Lake is ~55 m below sea-level, consisting of a marine-concentrated hypersaline brine about ten times seawater concentration (55). **(B)** Aerial photograph of Deep Lake, November 2008. **(C)** Sampling from the deepest point of Deep Lake. **(D)** Expedition work site at Deep Lake, November 2008, showing mobile work shelters and equipment for sampling. **(E)** View of Deep Lake from surrounding hills. Work site just visible near the shore on the right hand side of the panorama. The flat line of land in the background marks the water level before Deep Lake was isolated from the sea. **(F)** Relics of a past marine ecosystem are evident from tube worms and shells scattered around the shoreline and hills, and more contemporary preservation of carcasses of penguins and seals are present near shore, possibly having been preserved by the salt and cold for 100s to 1000s of years.


Crenarchaeota
non–halo Euryarchaeota
Nanohaloarchaea
Halomicrobium
Haloarcula
Halobacterium noricense, AJ548827.1
Halobacterium DL1, ARB_19876C8E
Halobacterium jilantaiense, AB477970.1
Halobacterium salinarum, AJ496185.1
Halococcus hamelinensis, NR_028168.1
Halobiforma
Haloterrigena/Natrinema
DEEP–7_14, AF142881.1
DL31, CP002988.1
DEEP–8_9, AF142882.1
DEEP–9_10, AF142883.1
Haloquadratum walsbyi, AM180088.1
Haloferax
Halogranum rubrum, FJ889134.1
Haloplanus natans, DQ417341.2
uncultured archaeon, CU467171.1
uncultured archaeon, CU467158.1
uncultured archaeon, DQ432517.1
uncultured archaeon, EU722677.1
tADL, JF421969.1
Halorubrum vacuolatum, NR_028224.1
Halorubrum orientalis, AM235786.1
Halorubrum halophilum, EF077637.1
Halorubrum saccharovorum, AM412370.1
Halorubrum lacusprofundi, 1U17365.
Halorubrum trapanicum, AB477979.1
Halorubrum alkaliphilum, AY510708.1
Halorubrum tibetense, AF435111.1
0.10

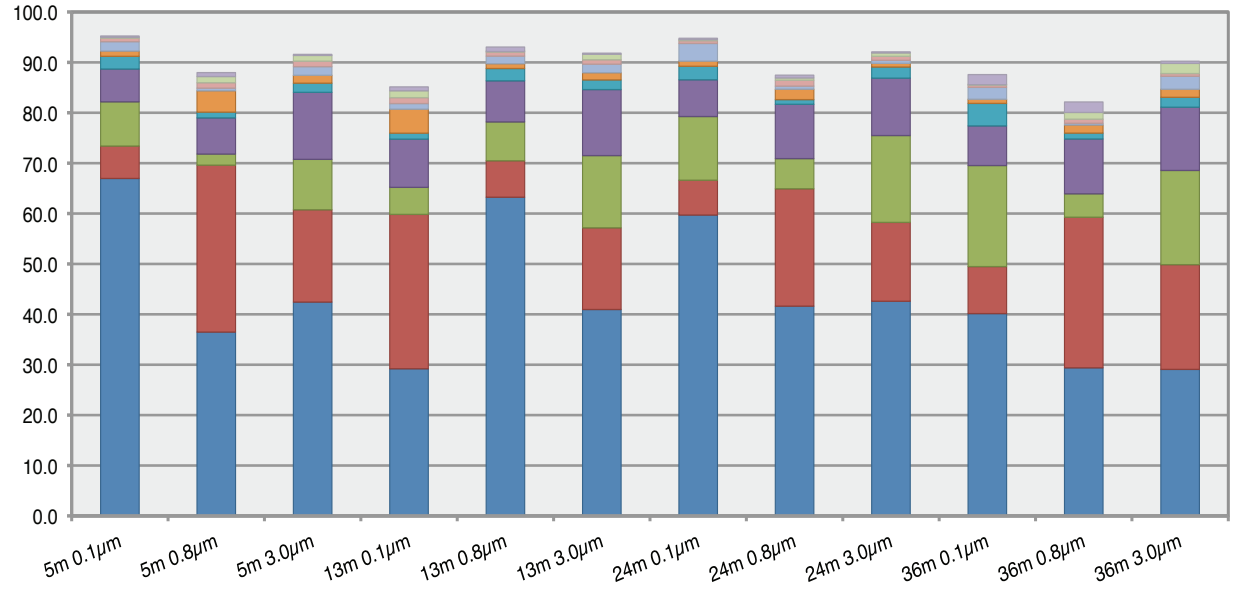

100.0
90.0
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
5m 0.1μm
5m 0.8μm
5m 3.0μm
13m 0.1μm
13m 0.8μm
13m 3.0μm
24m 0.1μm
24m 0.8μm
24m 3.0μm
36m 0.1μm
36m 0.8μm
36m 3.0μm

Fig. S2. Taxonomy and abundance of tADL, DL31, *Hl* and DL1 in Deep Lake. Upper panel: Neighbour-joining phylogenetic tree highlighting the relationships between the four DL haloarchaea tADL, DL31, *Hl* and DL1. All four DL isolates belong to distinct genera level taxa within the family *Halobacteriaceae*. Lower panel: Bar chart displaying the best taxonomic affiliation down to species level of SSU rRNA gene V6 tag sequences comprising >0.7% of all sequences derived from 5 m, 13 m , 24 m and 36 m depths in DL. tADL (blue), DL31 (red), *Dunaliella* chloroplast 100 (green), *Hl* (purple), DL29-clone (blue green), *Halospina denitrificans* (orange), *Chlorophyceae* (light blue), *Halobacteriaceae* (pink), *Oceanospirillales* (light green), *Bacteriodetes* (light purple).

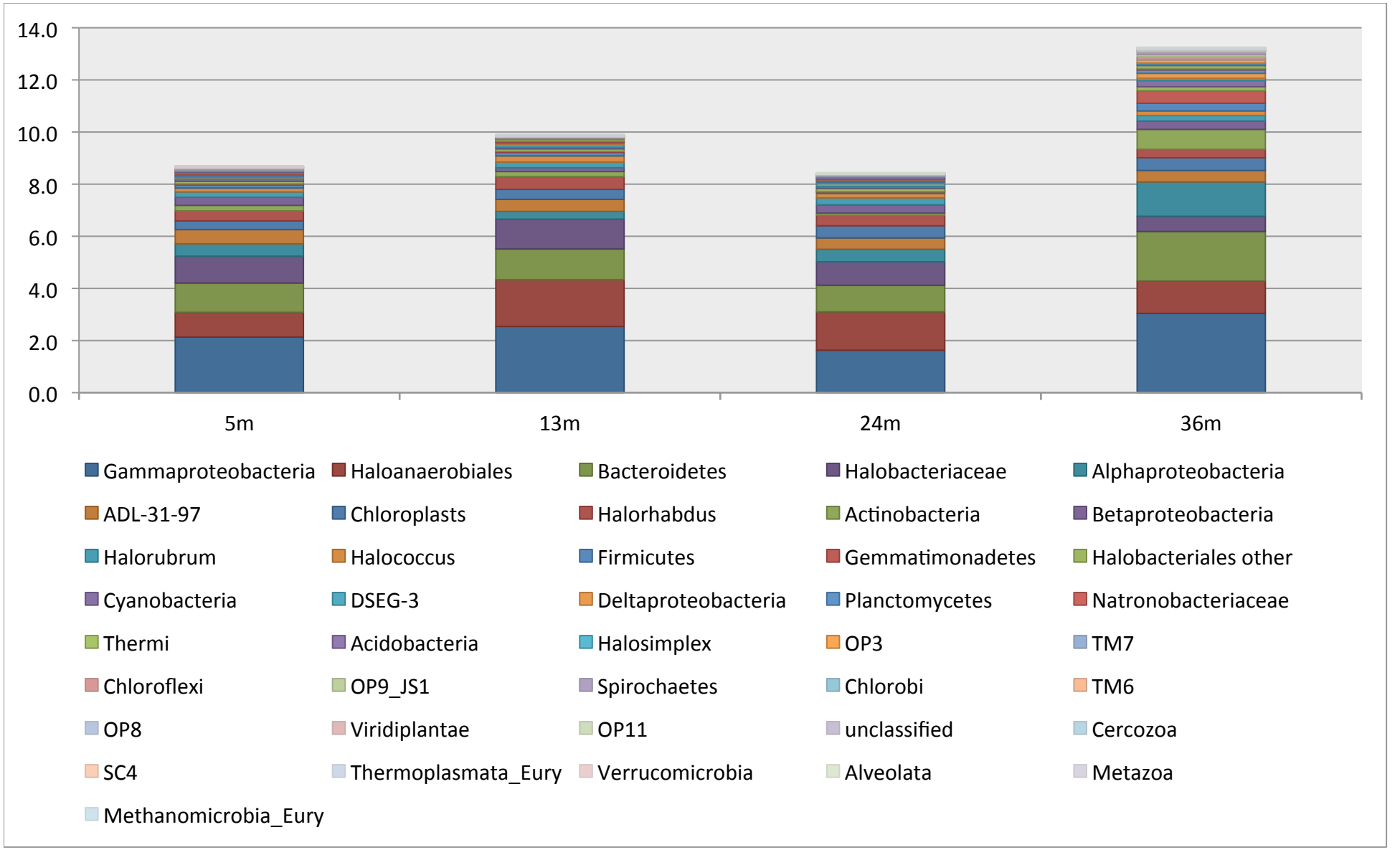

14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
5m
13m
24m
36m
Gammaproteobacteria
Haloanaerobiales
Bacteroidetes
Halobacteriaceae
Alphaproteobacteria
ADL-31-97
Chloroplasts
Halorhabdus
Actinobacteria
Betaproteobacteria
Halorubrum
Halococcus
Firmicutes
Gemmatimonadetes
Halobacteriales other
Cyanobacteria
DSEG-3
Deltaproteobacteria
Planctomycetes
Natronobacteriaceae
Thermi
Acidobacteria
Halosimplex
OP3
TM7
Chloroflexi
OP9_JS1
Spirochaetes
Chlorobi
TM6
OP8
Viridiplantae
OP11
unclassified
Cercozoa
SC4
Thermoplasmata_Eury
Verrucomicrobia
Alveolata
Metazoa
Methanomicrobia_Eury

Fig. S3. Taxonomy and abundance of rare species in Deep Lake. Bar chart displaying taxonomic affiliations of SSU rRNA gene V6 tag sequences comprising <0.7% of all sequences derived from 5 m, 13 m, 24 m and 36 m depths in DL. Affiliation has been presented at the family level, except in the case of *Proteobacteria* and *Halobacteriales* which have been further delineated (*e.g.* class). These data represent the taxa not already shown in Fig. S2.

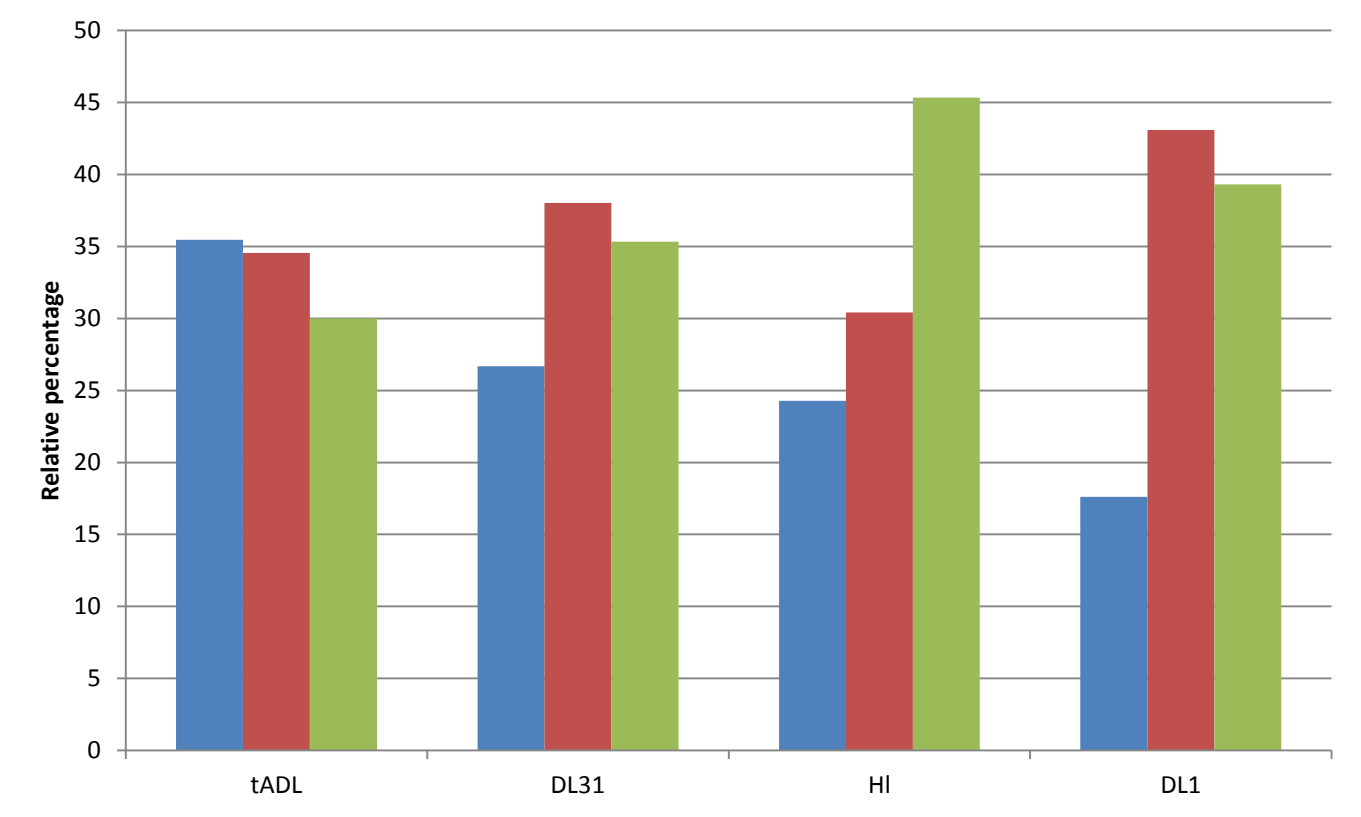

50
45
40
35
30
25
20
15
10
5
0
Relative percentage
tADL
DL31
Hl
DL1

Fig. S4. Relative abundance of tADL, DL31, *Hl* and DL1 genomes across Deep Lake. The relative abundance of the four DL genomes averaged across the three size fractions was determined from three analysis methods: SSU rRNA gene pyrotag sequencing (green bars) and FR from metagenomic data generated by Roche 454 Titanium (blue bars) and Solexa Illumina sequencing (red bars). tADL was the most abundant organism in all cases, followed by DL31 (2^{nd}), *Hl* (3^{rd}) and DL1 (4^{th}).

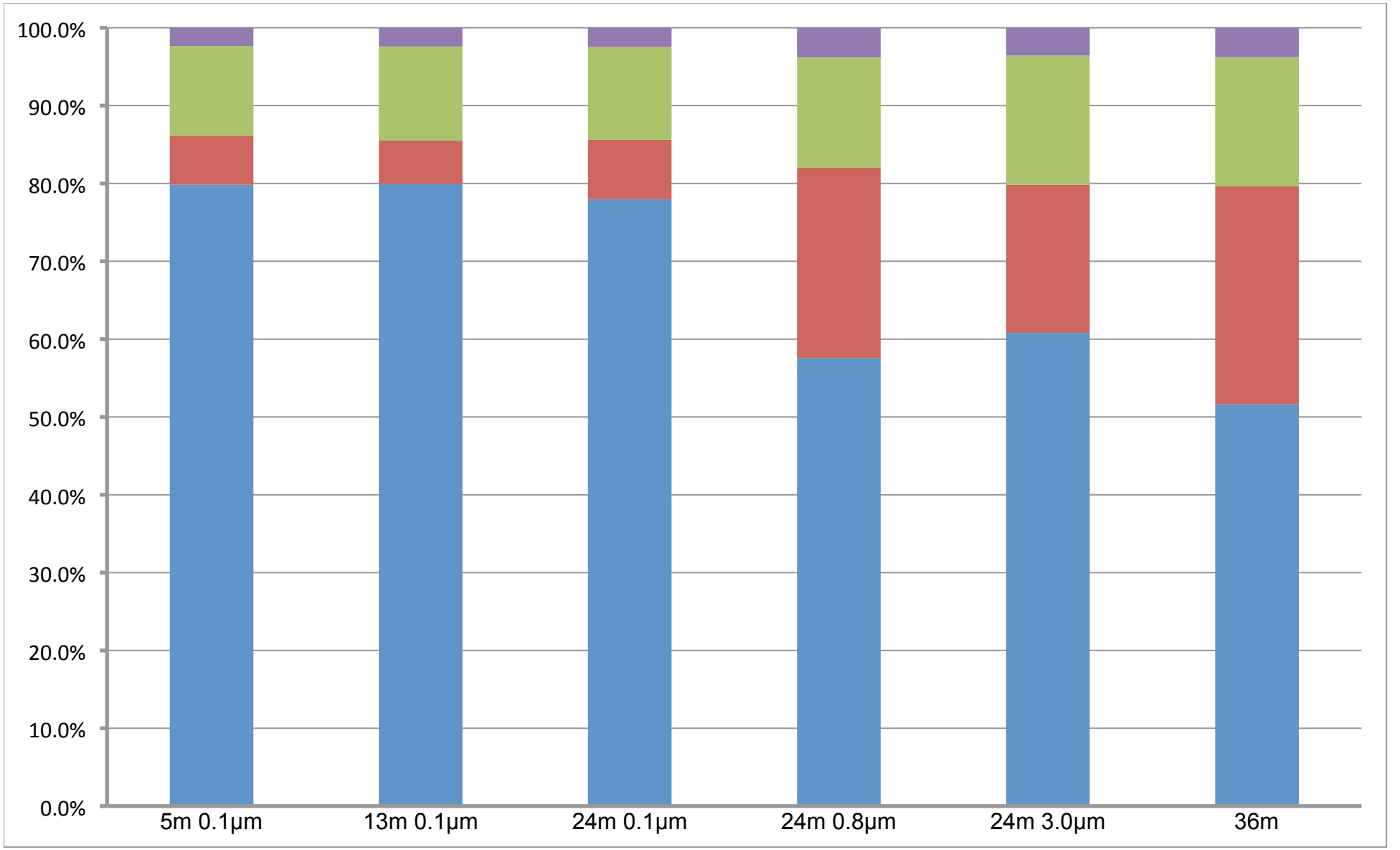
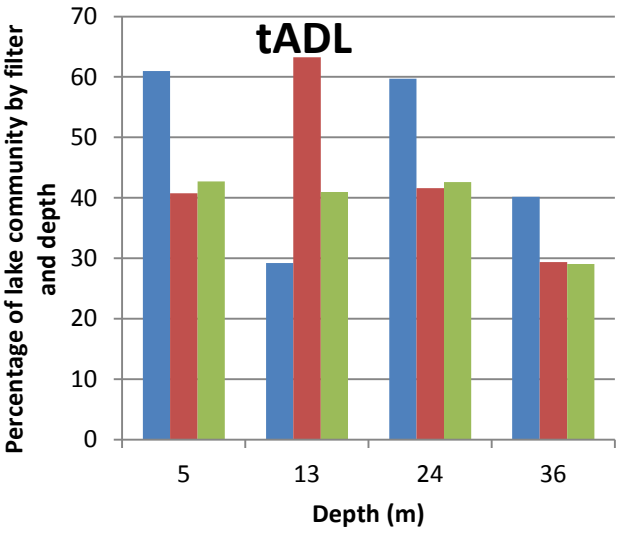
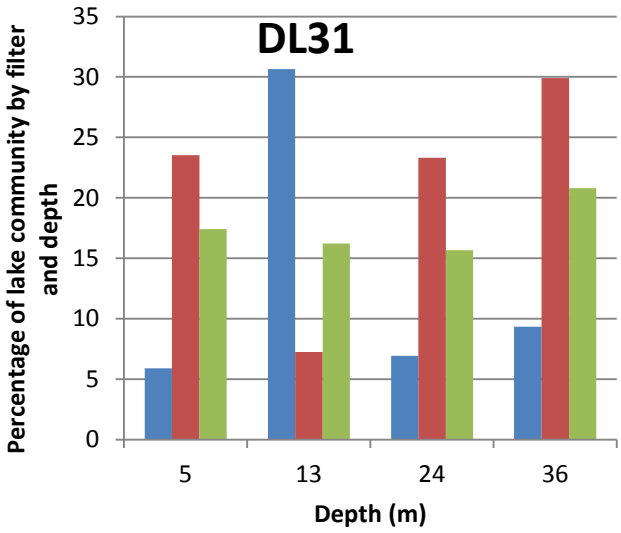
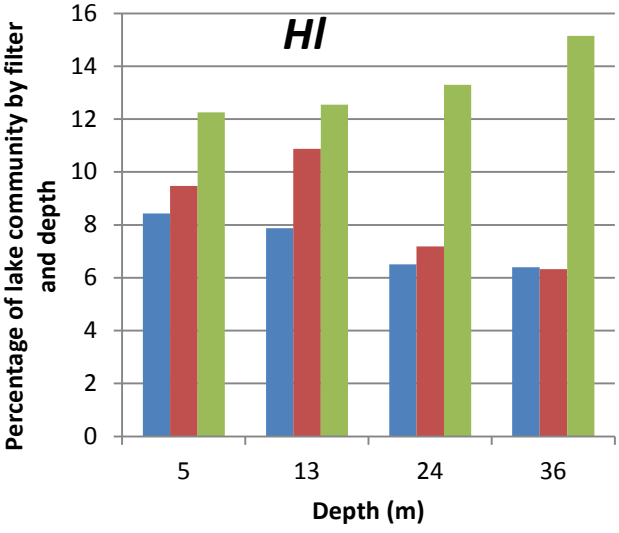
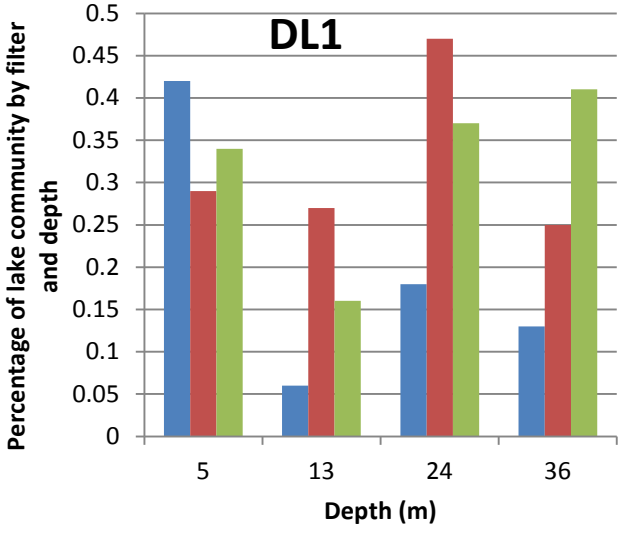

100.0%
90.0%
80.0%
70.0%
60.0%
50.0%
40.0%
30.0%
20.0%
10.0%
0.0%
5m 0.1µm
13m 0.1µm
24m 0.1µm
24m 0.8µm
24m 3.0µm
36m

Fig. S5. Fragment recruitment across depths and filter sizes. Relative abundance of fragments recruited to the four DL genomes (tADL: blue, DL31: red, *Hl*: green, DL1: purple) from 454 Titanium generated metagenomic reads for six dataset (5 m 0.1 µm; 13 m 0.1 µm; 24 m 0.1, 0.8 and 3.0 µm; 36m pooled).


tADL
Percentage of lake community by filter and depth
70
60
50
40
30
20
10
0
5
13
24
36
Depth (m)


DL31
Percentage of lake community by filter and depth
35
30
25
20
15
10
5
0
5
13
24
36
Depth (m)


Hl
Percentage of lake community by filter and depth
16
14
12
10
8
6
4
2
0
5
13
24
36
Depth (m)


DL1
Percentage of lake community by filter and depth
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
5
13
24
36
Depth (m)

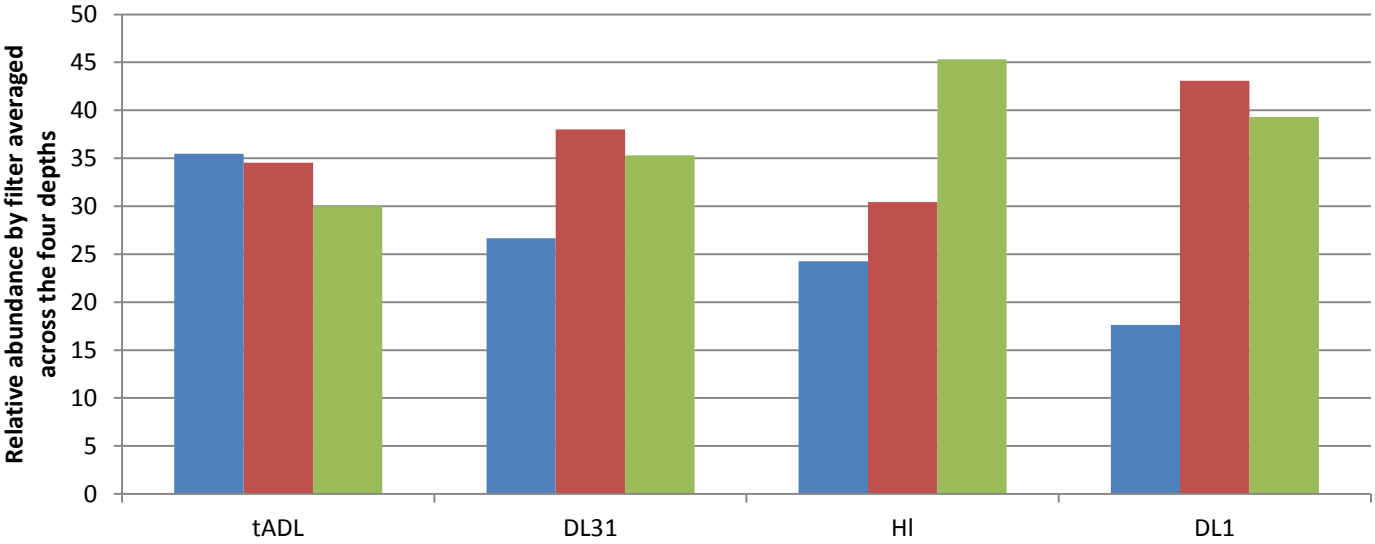

Relative abundance by filter averaged across the four depths
50
45
40
35
30
25
20
15
10
5
0
tADL
DL31
Hl
DL1

Fig. S6. Relative abundance of tADL, DL31, *Hl* and DL1 SSU rRNA genes across Deep Lake. Top four panels: Percentage of total SSU rRNA gene V6 tag sequences for tADL, DL31, *Hl* and DL1 for each filter size and each lake depth. Lower panel: Percentage of total SSU rRNA gene V6 tag sequences for tADL, DL31, *Hl* and DL1 for each filter size averaged across the four lake depths. 0.1 µm (blue bars), 0.8 µm (red bars), 3.0 µm (green bars).

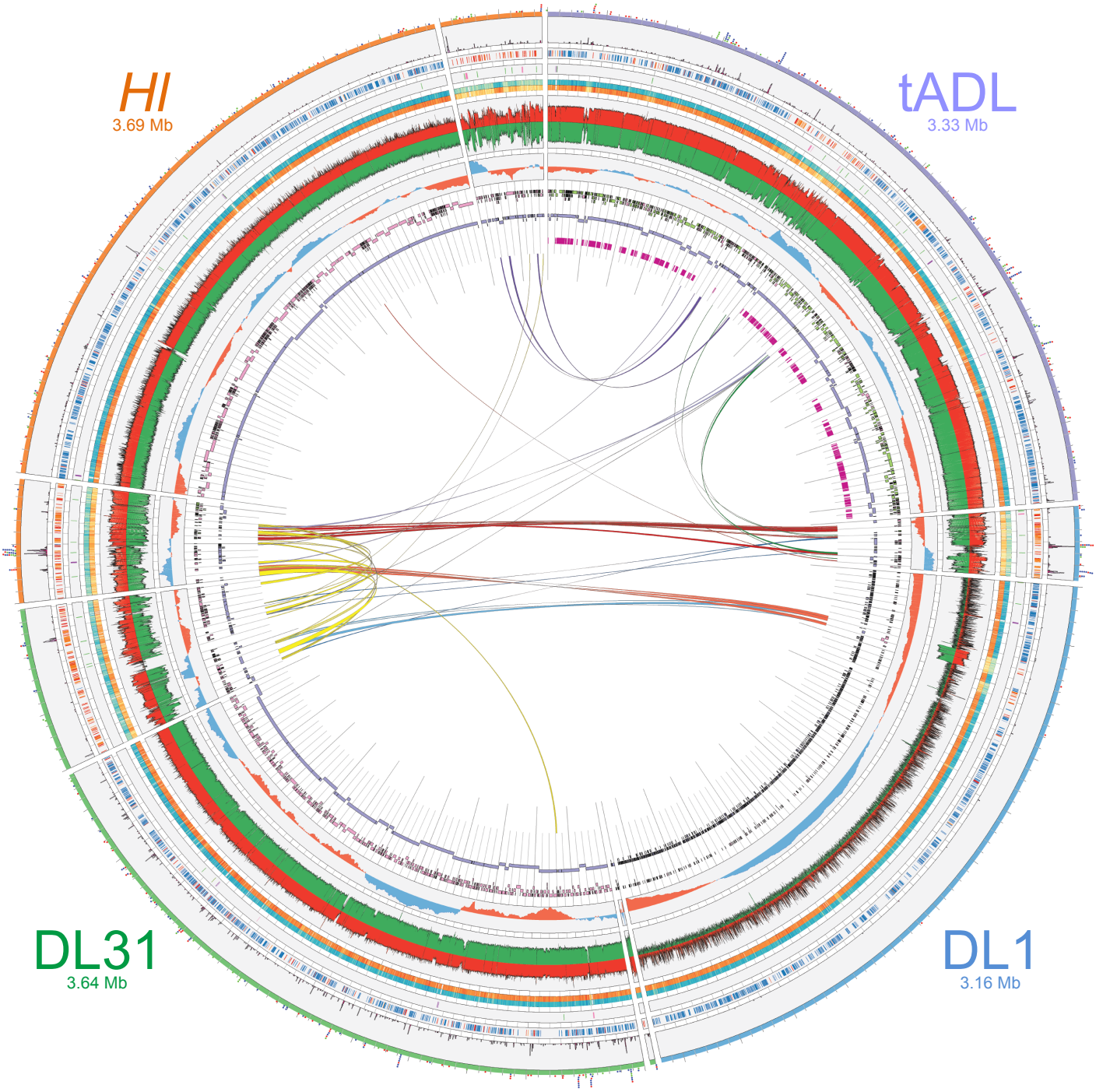

HI
3.69 Mb
tADL
3.33 Mb
DL31
3.64 Mb
DL1
3.16 Mb

Fig. S7. Genome characteristics of Deep Lake haloarchaea: tADL, DL1, DL31 and *Hl*. Circos plot (56), outside to inside: Outer filled dots: fixed SNPs (freq ≥ 0.9), non-synonymous (red), intergenic (green), synonymous (blue); 1st annulus: replicon backbones clockwise from the top, tADL (contig-32, purple), DL1 (contig-37, contig-38, light blue), DL31 (contig-113, contig-114, contig-115, green), *Hl* (NC_012028, NC_012029, NC_012030, orange); 2nd annulus: SNP histograms, stacked freq bands 0.7-0.8, 0.8-0.9, 0.9-1.0; 3rd annulus: core (blue), flexible (orange), IS (red); 4th annulus: CRISPR associated (pink), *orc1/cdc6* (green), rRNA (purple); 5th annulus: CAI (yellow-blue), CBI (yellow-orange) heatmaps, deeper colour indicates more adapted; 6th annulus: read-depth by gsMapper reference mapping (red), FR-hit fragment recruitment (green), log scale y-axis; 7th annulus: GC skew, > 0 (blue), < 0 (red), y-axis is independent per replicon; 8th, 9th annulus: Contigs from Celera WGS or gsMapper *de novo* assembly of 8 million 454 reads; 10th annulus: "ADL-like 5th genome" fragments aligned to the tADL genome; Internal lines: shared HIR of ≥ 99.8% nucleotide identity and 5 kb length.

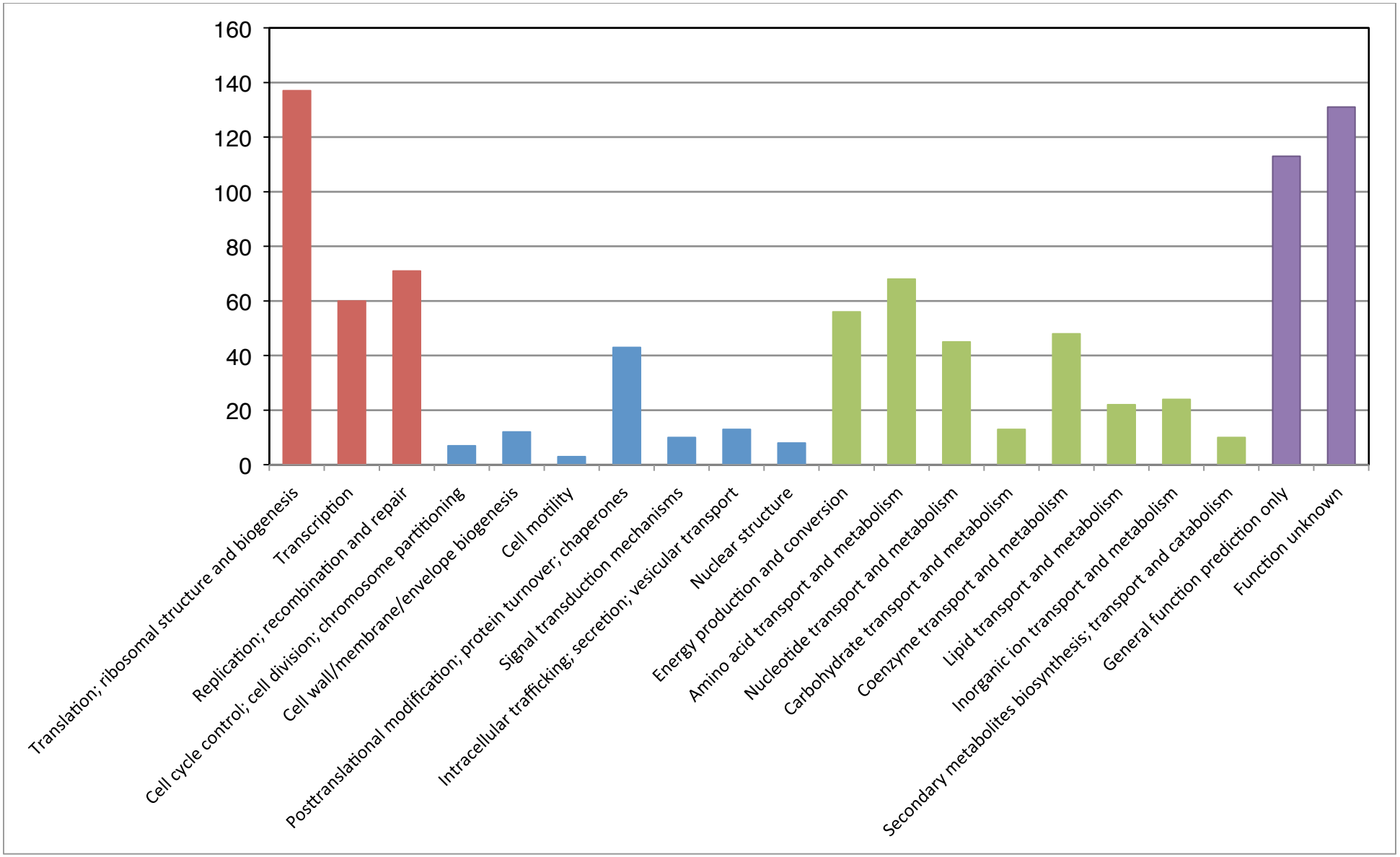

160
140
120
100
80
60
40
20
0
Translation; ribosomal structure and biogenesis
Transcription
Replication; recombination and repair
Cell cycle control; cell division; chromosome partitioning
Cell wall/membrane/envelope biogenesis
Cell motility
Posttranslational modification; protein turnover; chaperones
Signal transduction mechanisms
Intracellular trafficking; secretion; vesicular transport
Nuclear structure
Energy production and conversion
Amino acid transport and metabolism
Nucleotide transport and metabolism
Carbohydrate transport and metabolism
Coenzyme transport and metabolism
Lipid transport and metabolism
Inorganic ion transport and metabolism
Secondary metabolites biosynthesis; transport and catabolism
General function prediction only
Function unknown

Fig. S8. ArCOG assignments of the core gene content. Core orthologous groups assigned to ArCOG functional categories. Bars are coloured by the four functional classes: Information storage and processing (red), Cellular processes and signalling (blue), Metabolism (green) and Poorly characterised (purple).

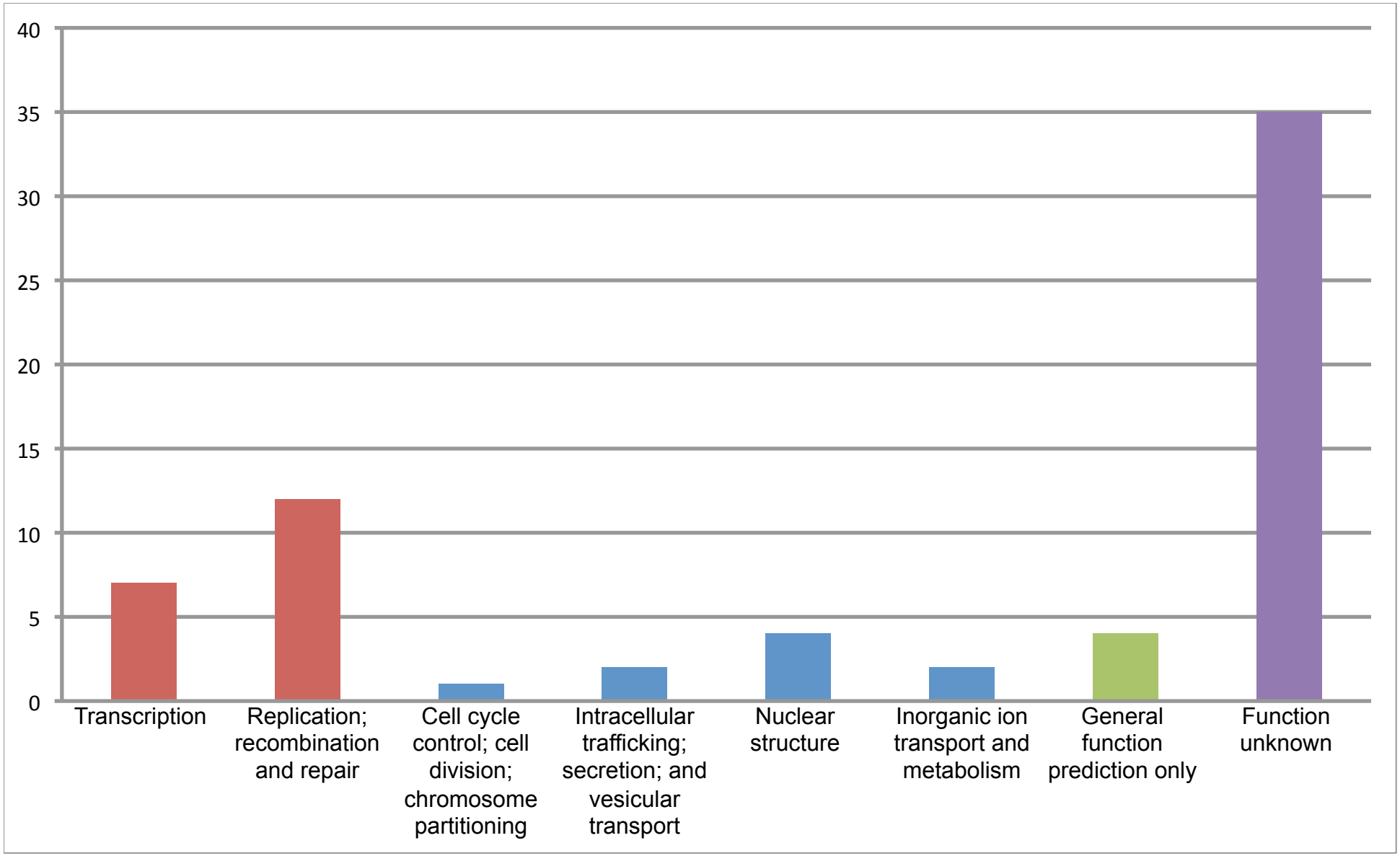

40
35
30
25
20
15
10
5
0
Transcription
Replication; recombination and repair
Cell cycle control; cell division; chromosome partitioning
Intracellular trafficking; secretion; and vesicular transport
Nuclear structure
Inorganic ion transport and metabolism
General function prediction only
Function unknown

Fig. S9. ArCOG assignments of the non-core gene content. Non-core orthologous groups assigned to ArCOG functional categories. Bars are coloured by the four functional classes: Information storage and processing (red), Cellular processes and signalling (blue), Metabolism (green) and Poorly characterised (purple).

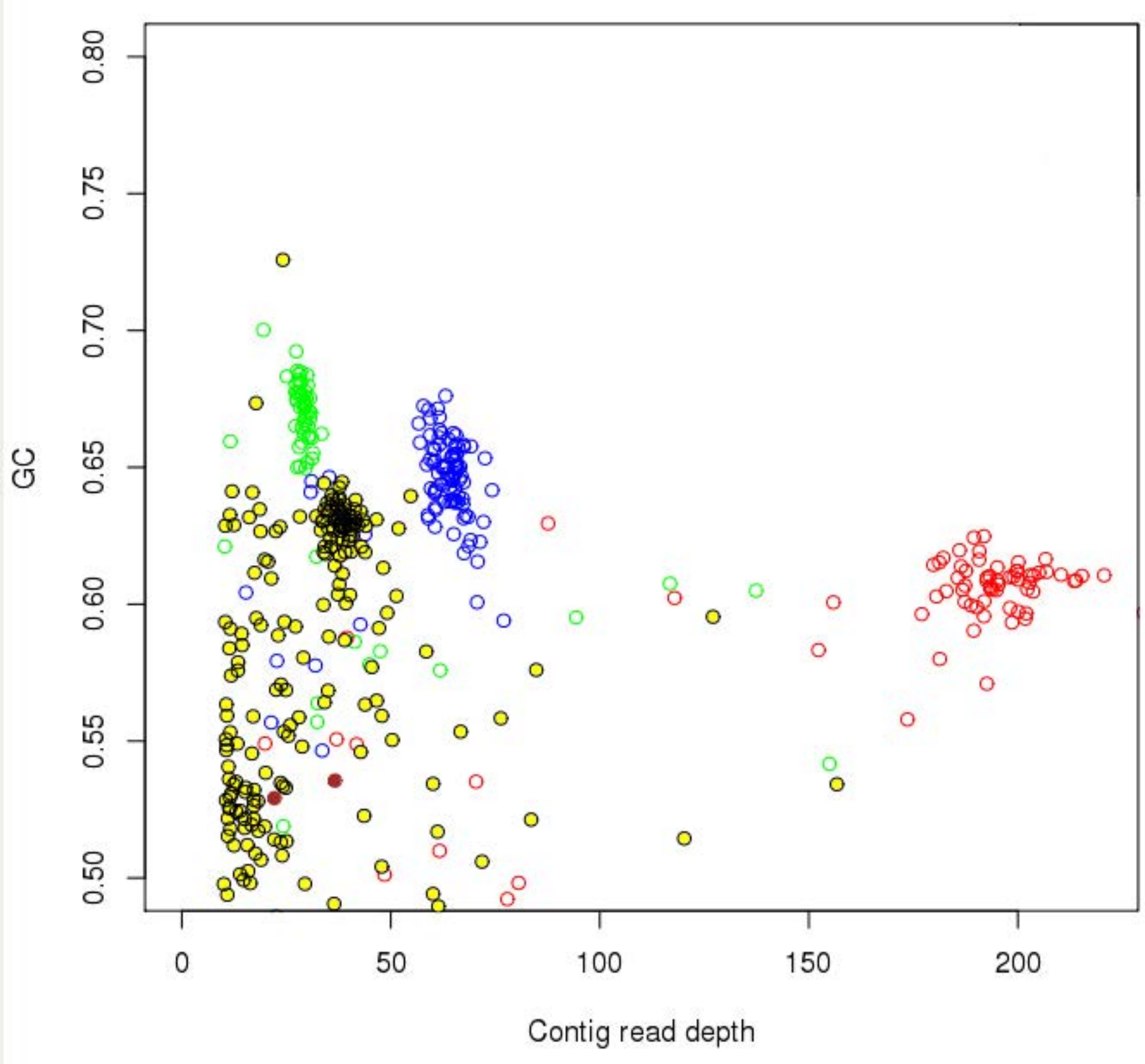

GC
0.50
0.55
0.60
0.65
0.70
0.75
0.80
0
50
100
150
200
Contig read depth

Fig. S10. Scatterplot of *de novo* assembly contigs in GC plus read-depth space. A 2-dimenional scatterplot of contigs longer than 15 kb from Celera WGS *de novo* assembly of all Roche 454 Titanium datasets, where axes are GC content and mean read-depth. Contigs possessing sufficient BLASTN scores (coverage > 90%, e-value < 10^{-10}) to the four DL genomes were labeled tADL (red), DL31 (blue), *Hl* (green), DL1 (brown). A dense cluster of unlabeled contigs (yellow) is visible with center (GC=0.63, RD=38.3) corresponding to the "tADL-like 5th genome".

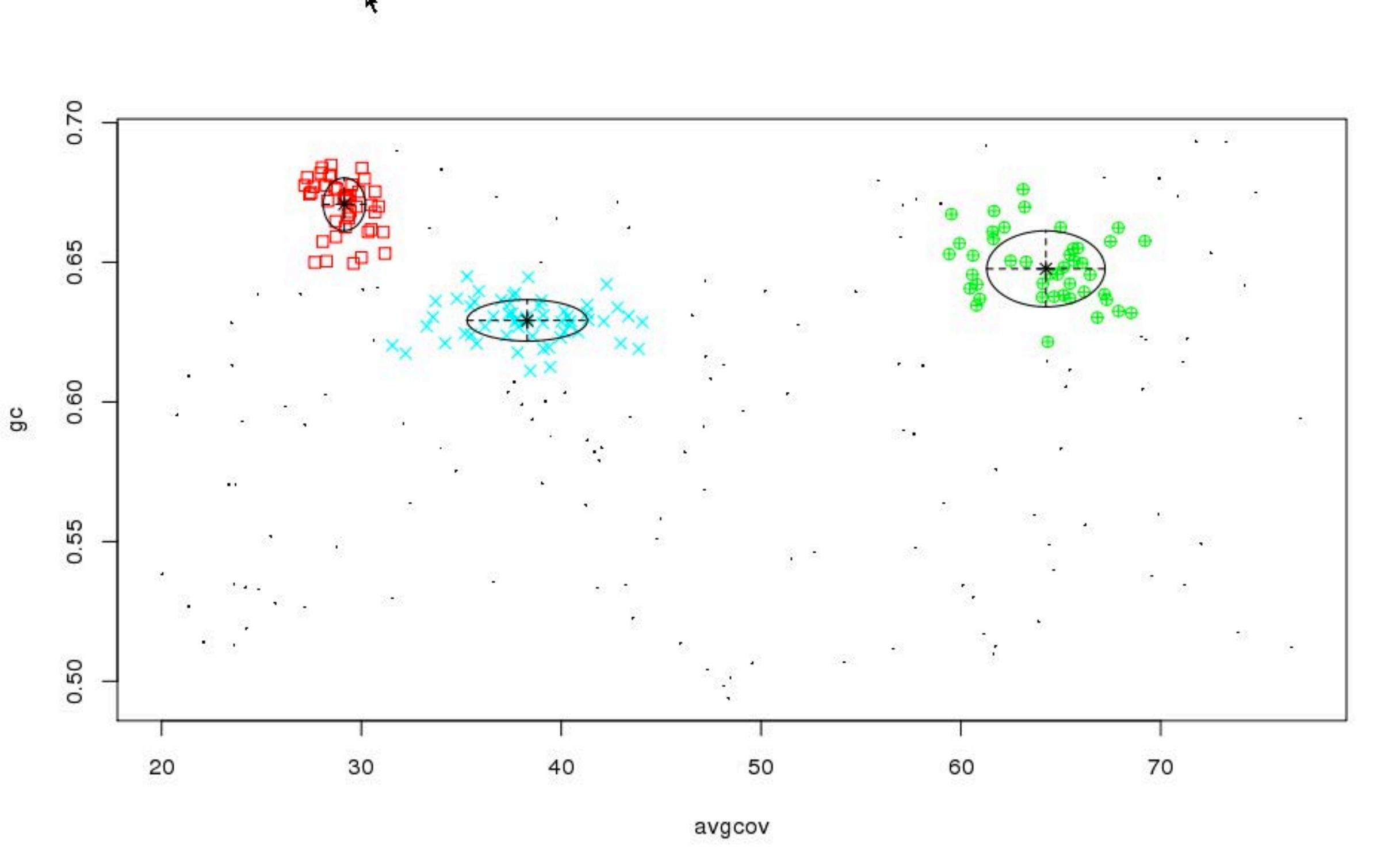

0.70
0.65
0.60
0.55
0.50
gc
20
30
40
50
60
70
avgcov

Fig. S11. Model based clustering of *de novo* assembly contigs in GC plus read-depth space. Contigs from Celera WGS *de novo* assembly of all Roche 454 Titanium samples were clustered by GC content and mean read-depth in R using Mclust model based clustering in the presence of Poisson noise (black specks). The space was limited to read-depths < 100. Three clusters were identified for DL31 (green circles), *Hl* (red squares) and the "tADL-related 5th genome" (cyan crosses).

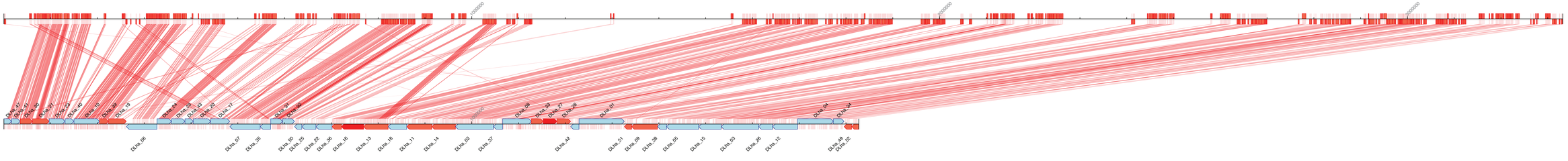

1000000
2000000
3000000
DLhk_47
DLhk_51
DLhk_30
DLhk_31
DLhk_23
DLhk_40
DLhk_10
DLhk_09
DLhk_19
DLhk_24
DLhk_39
DLhk_03
DLhk_20
DLhk_17
DLhk_01
DLhk_28
DLhk_06
DLhk_29
DLhk_27
DLhk_26
DLhk_07
DLhk_04
DLhk_34
DLhk_06
DLhk_07
DLhk_55
DLhk_40
DLhk_25
DLhk_22
DLhk_36
DLhk_16
DLhk_13
DLhk_18
DLhk_11
DLhk_14
DLhk_02
DLhk_37
DLhk_42
DLhk_51
DLhk_09
DLhk_38
DLhk_05
DLhk_15
DLhk_03
DLhk_26
DLhk_12
DLhk_49
DLhk_52

Fig. S12. Synteny plot of “tADL-related 5th genome” mapped to tADL. A synteny plot between the 52 contigs obtained from the *de novo* assembly identified as “tADL-related 5th genome”, and the tADL primary replicon was generated using CONTIGuator. Top line, tADL primary replicon; bottom line, “tADL-related 5th genome”. Overlapping contigs: light red, one side, dark red, both sides, blue, no overlap. Strandedness indicated by contig arrow.

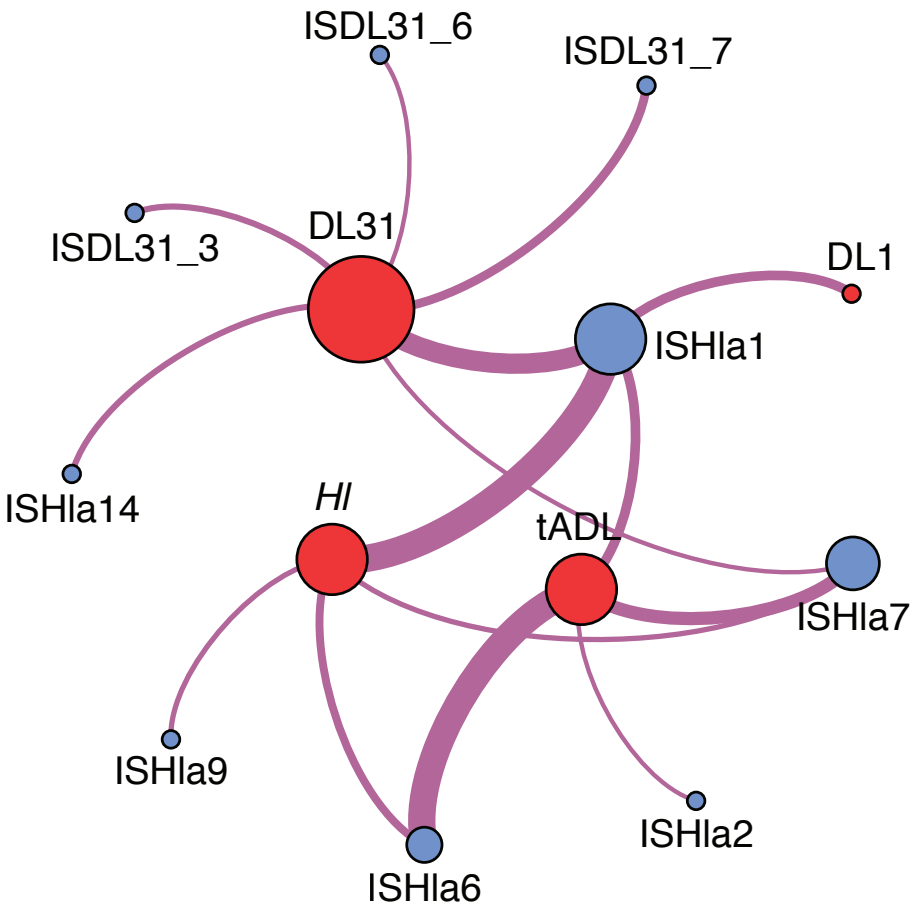

ISDL31_6
ISDL31_7
ISDL31_3
DL31
DL1
ISHla1
ISHla14
Hl
tADL
ISHla7
ISHla9
ISHla2
ISHla6

Fig. S13. Association networks for Deep Lake ISs. Bipartite association networks for ISs identified within the four DL haloarchaeal genomes, using Fruchterman-Reingold layout. ISs (blue nodes) and their containing genome (red nodes) where node radius scales linearly with weighted degree, and edge weights are proportional to frequency of IS occurrence.

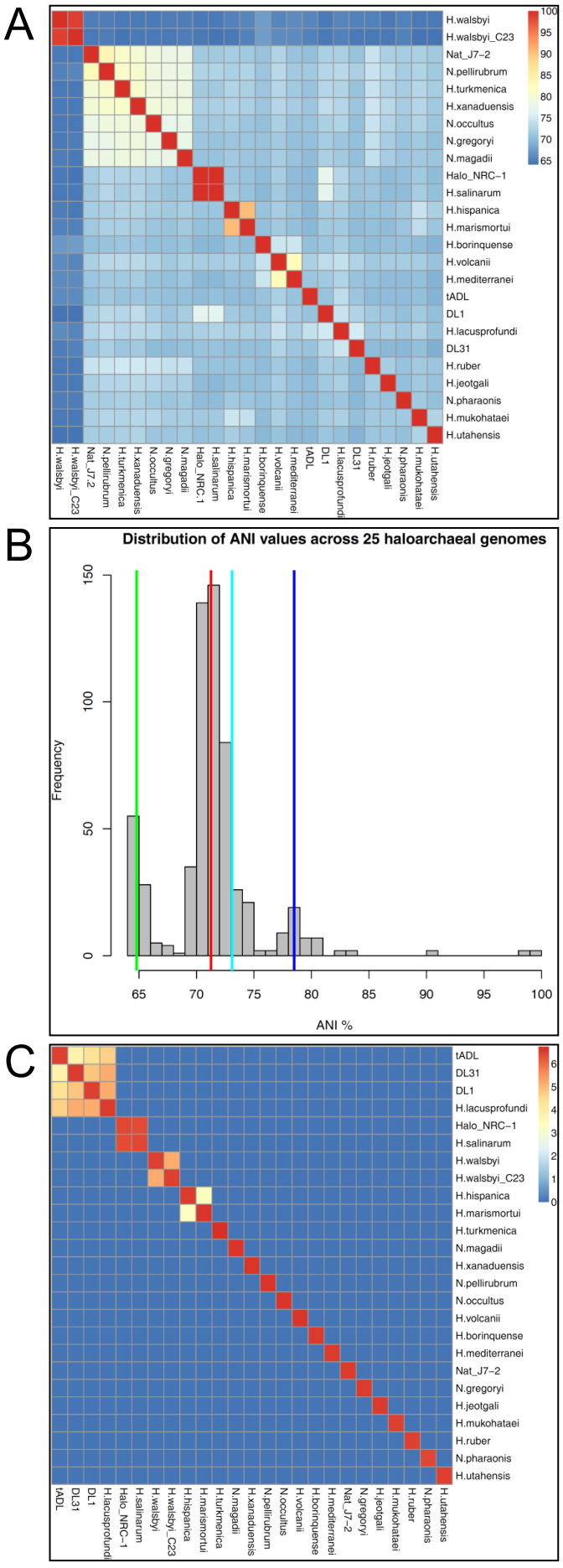

A
H.walsbyi
H.walsbyi_C23
Nat_J7~2
N.pellirubrum
H.turkmenica
H.xanaduensis
N.occultus
N.gregoryi
N.magadii
Halo_NRC~1
H.salinarum
H.hispanica
H.marismortui
H.borinquense
H.volcanii
H.mediterranei
tADL
DL1
H.lacusprofundi
DL31
H.ruber
H.jeotgali
N.pharaonis
H.mukohataei
H.utahensis
100
95
90
85
80
75
70
65
B
Distribution of ANI values across 25 haloarchaeal genomes
Frequency
0
50
100
150
65
70
75
80
85
90
95
100
ANI %
C
tADL
DL31
DL1
H.lacusprofundi
Halo_NRC~1
H.salinarum
H.walsbyi
H.walsbyi_C23
H.hispanica
H.marismortui
H.turkmenica
N.magadii
H.xanaduensis
N.pellirubrum
N.occultus
H.volcanii
H.borinquense
H.mediterranei
Nat_J7~2
N.gregoryi
H.jeotgali
H.mukohataei
H.ruber
N.pharaonis
H.utahensis
6
5
4
3
2
1
0

Fig. S14. ANI and HIR analysis of haloarchaeal genomes. Heatmap with rows and columns ordered by minimum variance for ANI (**A**), histogram of the distribution of ANI values (**B**), and heatmap with rows and columns ordered by minimum variance for extent of shared HIR (>99%, >2 kb) (**C**) between 25 completed haloarchaeal genomes. (**B**) median global ANI (red line); median ANI for Nat_J7-2, H.turkmenica, H.xanaduensis, N.pellirubrum, N.occultus, N.gregoryi, N.magadii (blue line); median ANI for H.walsbyi strains (green line); median ANI for DL haloarchaea, tADL, DL31, *Hl* and DL1 (light blue line).

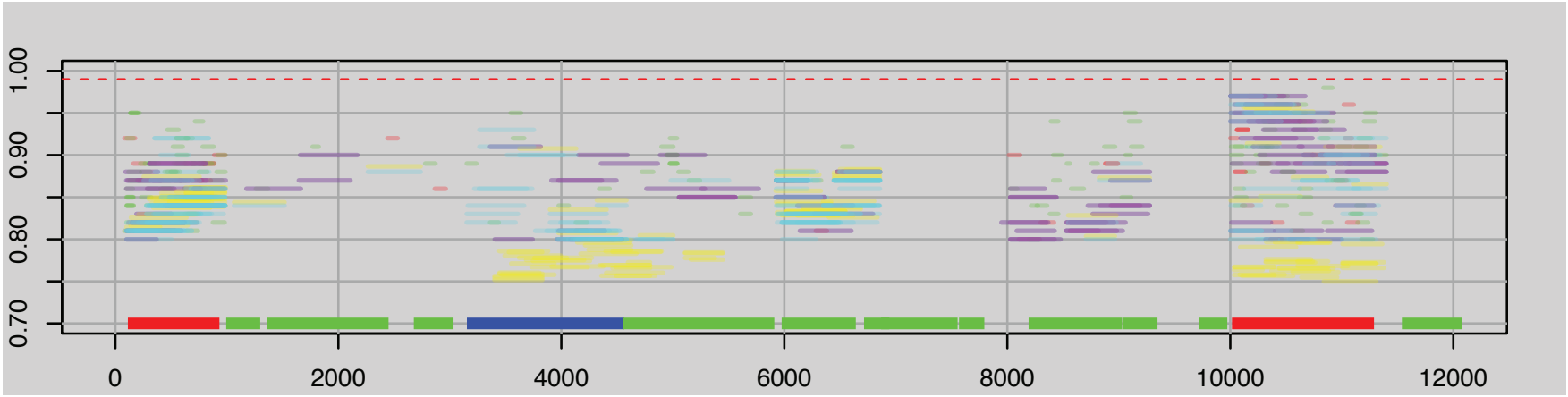

1.00
0.90
0.80
0.70
0
2000
4000
6000
8000
10000
12000

Fig. S15. HIR FR typical of metagenome data with matches from Chula Bay or Santa Pola. FR plot of 758 combined reads above 70% identity, 50% read coverage mapped from 6 high salinity saltern metagenomes to an HIR (13.3 kb, 99.9% identity) shared between DL31:Contig114 (664513..677829) and Hl:NC_012028 C (329458..342775). Reads are coloured by metagenome source: Chula Bay SRR023631 (red), SRR027043 (green) and Santa Pola SRR062267 (blue), SRR316684 (yellow), SRR328982 (magenta), SRR328983 (light blue). The red dashed line indicates 99% identity threshold, which no read exceeded and the thick line segments at the bottom represent genes colored by broad annotation categories: transposase (red), hypothetical (green), non-hypothetical (blue). Reads tend to pile up predominately on IS associated genes such as transposases. A bias toward FR in coding regions and away from intergenic regions is also apparent.

Table S1. Metagenome sequencing summary.

Library name	Technology	Depth (m)	Filter size (µm)	Raw reads
HTAA	454 Titanium	5	0.1	1,007,472
HTAF	454 Titanium	13	0.1	1,158,857
HTAC	454 Titanium	24	0.1	1,108,448
HTAB	454 Titanium	24	0.8	1,063,876
GWTB	454 Titanium	24	3.0	931,207
HTSY	454 Titanium	36	pooled	1,363,684
HWGG	454 Titanium	36	pooled	1,423,784
GXFW	Solexa Illumina	24	3.0	45,601,760

Table S2. Genome characteristics of tADL, DL31, *Hl* and DL1.

Name	Year of isolation	Pseudonym in Refseq	Replicon Count[1]	Size (bp)	Genes	16S rRNA genes	Coding genes	GC (%)	Coding bases (%)
tADL	water sample Dec 2006, isolation 2007	halophilic archaeon True-ADL	1	3332022	3520	2	3465	58.9	85.9
Hl	date of water sampling not stated – paper published in 1988	Halorubrum lacusprofundi	3	3692576	3725	3	3665	64	84.5
DL31	water sample Dec 2006, isolation 2009	halophilic archaeon DL31	3	3643158	3788	2	3737	62.4	83.2
DL1	water sample Dec 2006, isolation 2009	Halobacterium sp. DL	2	3162560	3367	1	3317	66.4	88.2

1: Replicons “assembled, autonomous circular genetic elements” were defined during the assembly process.

Table S3. SSU rRNA gene similarity matrix for tADL, DL31, *Hl* and DL1.

Dissimilarity	tADL	DL31	*Hl*	DL1		**Similarity**	tADL	DL31	*Hl*	DL1
tADL	0	0.1535	0.1347	0.171		tADL	1	0.8465	0.8653	0.829
DL31	0.1535	0	0.1559	0.1563		DL31	0.8465	1	0.8441	0.8437
Hl	0.1347	0.1559	0	0.1438		*Hl*	0.8653	0.8441	1	0.8562
DL1	0.171	0.1563	0.1438	0		DL1	0.829	0.8437	0.8562	1

Table S4. Abundance of taxa calculated from SSU rRNA gene pyrotag sequencing data.

Organism	5 m	13 m	24 m	36 m	Average
tADL	48.62	44.47	47.97	32.87	43.48
DL31	19.29	18.03	15.29	20.01	18.16
Dunaliella chloroplast100	7.00	9.13	11.95	14.46	10.64
Halorubrum lacusprofundi	8.99	10.28	9.82	10.43	9.88
DL29-clone	1.82	1.84	1.93	2.53	2.03
Halospina denitrificans	2.29	2.39	1.28	1.35	1.83
Dunaliella-18S	1.36	1.44	1.57	1.75	1.53
Halobacteriacea_cloneGX3	0.96	0.96	0.84	0.60	0.84
Halomonas subglaciescola	0.84	0.85	0.44	1.15	0.82
Bacterioidetes-Sphingobacteriales-ELB25-178	0.37	0.59	0.31	1.50	0.69
Natronoarchaeaum mannanilyticum	0.69	0.74	0.61	0.42	0.62
Gammaproteobacteria_Ectothiorhodospiraceae_cloneSINI729	0.55	0.71	0.42	0.70	0.60
Gammaproteobacteria_Salinisphaera_sp	0.48	0.99	0.52	0.35	0.59
Firmicute-Haloanaerobium_cloneARDBACWH2	0.33	0.84	0.61	0.47	0.56
Dunaliella chloroplat 97	0.33	0.38	0.47	0.49	0.42
Firmicute-Haloanaerobium	0.22	0.32	0.31	0.38	0.31
DL1	0.34	0.41	0.29	0.16	0.30
Psychroflexus torquis	0.26	0.24	0.15	0.46	0.28
Firmicute-Haloanaerobacter_lacunarum	0.28	0.42	0.33	0.08	0.28
ADL31_97	0.30	0.24	0.27	0.27	0.27
Other	*4.67*	*4.73*	*4.61*	*9.57*	*5.90*

[1] Relative abundance calculated from SSU rRNA gene pyrotag sequencing data of the 20 most abundant operational taxonomic units across all depths in DL.

Table S5. Insertion sequences in tADL, DL31, *Hl* and DL1 identified in the ISsaga database.

			ISFinder search[2]				Frequency[3]				
IS name	**Template ORF**	**IS Family**	**IS Group**	**Closest relative**	**Identity**	**Bitscore**	**tADL**	**DL1**	**DL31**	***Hl***	**Total**
ISHla1	orf00115	ISH3		H.lacus ISHla1	99	2752	11	11	23	31	76
ISHla6	orf00119	IS5		H.lacus ISHla6	100	2119	31	0	1	9	41
ISHla7	orf00143	ISNYC	ISH6	H.salin ISH6	88	1505	15	3	5	6	29
ISDL31_7	orf00149	IS66	ISBst12	H.wals ISHwa10	86	1043	0	2	10	0	12
ISHlac14	orf00153	ISH3		H.salin ISH20A	90	1493	0	1	7	3	11
ISHla2	orf00157	IS5	ISH1	H.lacus ISHla2	100	2365	5	1	1	3	10
ISHla8	orf00207	IS4	ISH8	H.sp NRC-1 ISH8B	92	1901	0	2	4	2	8
ISDL31_3	orf_ISf_6	IS6		H.sp NRC-1 ISH29	82	260	0	0	7	0	7
ISHla9	orf00175	ISH3		H.wals ISHwa13	81	509	0	0	1	6	7
ISDL31_6	orf00668	IS66	ISBst12	H.wals ISHwa10	86	932	0	0	5	0	5
ISDL31_2p	orf00872	-					0	0	4	0	4
ISHlac11	orf00776	IS200/IS605	IS1341	N.phara ISNph17	98	92	1	0	0	3	4
ISHlac15	orf_ISf_5	IS6		N.phara ISNph1	96	50	0	0	1	3	4
ISHlac10	orf00355	IS200/IS605	IS1341	N.phara ISNph18	97	62	0	0	0	3	3
ISDL31_11	orf00661	IS200/IS605		H.sp NRC-1 ISH12	82	149	0	1	0	2	3
ISNph18	orf02643	IS200/IS605	IS1341	N.phara ISNph18	93	50	0	0	3	0	3
ISDL31_9	orf00945	IS5	ISH1	H.utah	91	1211	0	3	0	0	3
ISHla3	orf01135	IS5	ISH1	H.lacus ISHla3	100	1842	0	0	1	2	3
ISDL31_10	orf01662	IS200/IS605		N.phara ISNph5	97	60	0	2	0	0	2
ISHlac12	orf00692	-					0	0	0	1	1
ISHlac13	orf00739	-					0	0	0	1	1
ISDL31_1	orf04615	IS1595	ISH4	H.sp ISH4	86	920	0	0	1	0	1
ISHla4	orf00615	IS5	ISH1	H.lacus ISHla4	100	3576	0	0	0	1	1
ISDL31_5	orf_ISf_1	IS6		H.maris ISH15	86	603	0	0	1	0	1
ISHla5	orf00210	ISH3		H.lacus ISHla5	100	2769	0	0	0	1	1
ISDL31_8_p	orf02573	ISNYC	ISH6	H.salin ISH6	87	1417	0	0	1	0	1

[1] Insertion sequences identified and manually annotated in ISsaga from the four DL genomes. ISs that have been previously identified (red); newly defined ISs given identifying names (IS name) following the advised nomenclature (black).
[2] ISFinder public database of ISs nearest relative identify is shown if the IS has been previously discovered (>95% similarity).
[3] Frequencies within each DL genome and total occurrence across the 4 DL genomes.

Table S6. Fixed SNPs in the genomes of tADL, DL31, HI and DL1. Fixed SNP (frequency > 0.9) for all replicons of the four DL genomes. For coding SNPs, ArCOG functional assignments are included from the containing gene.

Replicon	Pos	Ref	amb Var	Var	Freq	CDS	SYN	Locus	A	C	G	T	Tot	dinuc	Locus2	GI	arAcc	arCOG	Cat	Class	COG	COG Class	Description
Contig113	4168	A	G	G	0.936	T	F	Halar_0007	33	0	484	0	517	ga	Halar_0007	2507077907	gi|345133538|ref|YP_004821329.1|	arCOG02413	4	R	COG1598	S	Predicted nuclease of the RNAse H fold, HicB family
Contig113	4264	C	A	A	0.948	T	F	Halar_0008	385	21	0	0	406	gc	Halar_0008	2507077908	gi|345133539|ref|YP_004821330.1|	arCOG03087	4	R	COG1724	N	Predicted RNA binding protein (dsRBD-like fold), HicA family
Contig113	4391	A	G	G	0.928	T	F	Halar_0008	29	1	385	0	415	ga	Halar_0008	2507077908	gi|345133539|ref|YP_004821330.1|	arCOG03087	4	R	COG1724	N	Predicted RNA binding protein (dsRBD-like fold), HicA family
Contig114	89550	A	C	C	0.946	F	F		12	210	0	0	222	ca									
Contig114	219460	C	T	T	0.931	F	F		0	18	0	242	260	gc									
Contig114	226214	A	C	C	0.922	F	F		17	200	0	0	217	ca									
Contig114	352167	G	T	T	0.955	T	T	Halar_0346	0	0	2	42	44	ag	Halar_0346	2507078246	gi|345007123|ref|YP_004809975.1|	arCOG02751	1	L	COG3039	L	Transposase, IS5 family
Contig114	352185	G	A	A	0.955	T	T	Halar_0346	42	0	2	0	44	gg	Halar_0346	2507078246	gi|345007123|ref|YP_004809975.1|	arCOG02751	1	L	COG3039	L	Transposase, IS5 family
Contig114	493035	T	C	C	0.995	T	F	Halar_0503	1	184	0	0	185	ct	Halar_0503	2507078403	gi|345007266|ref|YP_004810118.1|	arCOG03948	1	K			DNA or RNA helicase of superfamily II
Contig114	505157	C	T	T	0.929	T	F	Halar_0510	0	8	0	105	113	ac									
Contig114	547848	C	T	T	0.915	T	T	Halar_0555	0	26	1	291	318	gc	Halar_0555	2507078455	gi|345007311|ref|YP_004810163.1|	arCOG06281	4	S			Uncharacterized conserved protein
Contig114	548319	A	C	C	0.917	T	T	Halar_0555	19	211	0	0	230	ga	Halar_0555	2507078455	gi|345007311|ref|YP_004810163.1|	arCOG06281	4	S			Uncharacterized conserved protein
Contig114	548346	T	C	C	0.92	T	T	Halar_0555	0	230	0	20	250	gt	Halar_0555	2507078455	gi|345007311|ref|YP_004810163.1|	arCOG06281	4	S			Uncharacterized conserved protein
Contig114	557624	A	G	G	0.951	T	T	Halar_0560	10	0	193	0	203	ga	Halar_0560	2507078460	gi|345007316|ref|YP_004810168.1|	arCOG06162	4	S			Cell suface protein
Contig114	628187	A	C	C	1	T	F	Halar_0633	0	49	0	0	49	aa	Halar_0633	2507078533	gi|345007382|ref|YP_004810234.1|	arCOG04525	4	S	COG3119	P	Uncharacterized conserved protein
Contig115	31546	C	T	T	0.932	F	F		0	18	0	248	266	cc									
Contig115	32896	A	G	G	0.902	T	T	Halar_0750	22	0	202	0	224	ca	Halar_0750	2507078651	gi|345004041|ref|YP_004806894.1|	arCOG04603	3	G	COG0158	G	Fructose-1,6-bisphosphatase
Contig115	64188	A	G	G	0.964	T	T	Halar_0783	7	0	186	0	193	ca	Halar_0783	2507078685	gi|345004074|ref|YP_004806927.1|	arCOG07564	4	S			Uncharacterized conserved protein
Contig115	64190	T	C	C	0.963	T	F	Halar_0783	0	184	0	7	191	at	Halar_0783	2507078685	gi|345004074|ref|YP_004806927.1|	arCOG07564	4	S			Uncharacterized conserved protein
Contig115	64358	A	G	G	0.926	F	F		19	0	237	0	256	ta									
Contig115	64385	C	T	T	0.974	F	F		0	4	0	148	152	ac									
Contig115	64751	T	C	C	0.98	T	T	Halar_0784	0	240	0	5	245	gt	Halar_0784	2507078686	gi|345004075|ref|YP_004806928.1|	arCOG02202	2	D	COG0206	D	Cell division GTPase
Contig115	65132	G	A	A	0.903	T	T	Halar_0784	28	0	3	0	31	cg	Halar_0784	2507078686	gi|345004075|ref|YP_004806928.1|	arCOG02202	2	D	COG0206	D	Cell division GTPase
Contig115	65148	G	A	A	0.955	T	F	Halar_0784	42	0	2	0	44	tg	Halar_0784	2507078686	gi|345004075|ref|YP_004806928.1|	arCOG02202	2	D	COG0206	D	Cell division GTPase
Contig115	65153	A	G	G	0.966	T	T	Halar_0784	2	0	57	0	59	ca	Halar_0784	2507078686	gi|345004075|ref|YP_004806928.1|	arCOG02202	2	D	COG0206	D	Cell division GTPase
Contig115	74344	A	G	G	0.907	T	T	Halar_0795	26	0	254	0	280	ca	Halar_0795	2507078697	gi|345004085|ref|YP_004806938.1|	arCOG00162	3	E	COG1177	E	ABC-type spermidine/putrescine transport system, permease component II
Contig115	109923	C	T	T	0.913	T	F	Halar_0830	0	7	0	73	80	cc	Halar_0830	2507078733	gi|345004120|ref|YP_004806973.1|	arCOG02708	3	E			Lactoylglutathione lyase or related enzyme
Contig115	120227	C	G	G	0.919	T	T	Halar_0842	0	18	204	0	222	ac	Halar_0842	2507078745	gi|345004130|ref|YP_004806983.1|	arCOG04232	3	I	COG1884	I	Methylmalonyl-CoA mutase
Contig115	182903	A	C	C	0.903	F	F		20	187	0	0	207	ta									
Contig115	224786	T	G	G	0.944	F	F		0	0	185	11	196	at									
Contig115	227468	A	G	G	0.995	T	T	Halar_0945	0	0	194	1	195	ca	Halar_0945	2507078850	gi|345004230|ref|YP_004807083.1|	arCOG01389	2	M			Glycosyltransferase
Contig115	231976	A	G	G	0.996	T	F	Halar_0949	1	0	238	0	239	ga	Halar_0949	2507078854	gi|345004234|ref|YP_004807087.1|	arCOG09197	4	R			SAM-dependent methyltransferase
Contig115	234154	C	T	T	0.978	T	T	Halar_0952	0	4	0	174	178	tc	Halar_0952	2507078857	gi|345004237|ref|YP_004807090.1|	arCOG01623	3	C	COG0667	C	Predicted oxidoreductase (related to aryl-alcohol dehydrogenase)
Contig115	234716	C	G	G	0.969	T	F	Halar_0952	0	3	94	0	97	ac	Halar_0952	2507078857	gi|345004237|ref|YP_004807090.1|	arCOG01623	3	C	COG0667	C	Predicted oxidoreductase (related to aryl-alcohol dehydrogenase)
Contig115	234908	A	G	G	0.921	T	F	Halar_0952	19	0	223	0	242	ta	Halar_0952	2507078857	gi|345004237|ref|YP_004807090.1|	arCOG01623	3	C	COG0667	C	Predicted oxidoreductase (related to aryl-alcohol dehydrogenase)
Contig115	234972	G	A	A	0.945	T	F	Halar_0952	222	0	13	0	235	gg	Halar_0952	2507078857	gi|345004237|ref|YP_004807090.1|	arCOG01623	3	C	COG0667	C	Predicted oxidoreductase (related to aryl-alcohol dehydrogenase)
Contig115	235054	A	C	C	0.92	T	T	Halar_0952	21	242	0	0	263	ga	Halar_0952	2507078857	gi|345004237|ref|YP_004807090.1|	arCOG01623	3	C	COG0667	C	Predicted oxidoreductase (related to aryl-alcohol dehydrogenase)
Contig115	235141	A	G	G	0.922	T	T	Halar_0952	19	0	225	0	244	ca	Halar_0952	2507078857	gi|345004237|ref|YP_004807090.1|	arCOG01623	3	C	COG0667	C	Predicted oxidoreductase (related to aryl-alcohol dehydrogenase)
Contig115	235193	T	C	C	0.914	F	F		0	265	0	25	290	tt									
Contig115	253962	A	T	T	0.951	F	F		17	0	0	333	350	ga									
Contig115	254037	T	C	C	0.945	F	F		1	309	0	17	327	at									

Contig115	265011	T	C	C	0.935	T	F	Halar_0979	1	201	0	13	215	gt	Halar_0979	2507078884	gi|345004264|ref|YP_004807117.1|	arCOG02267	3	P			Phosphate/sulphate permease
Contig115	281311	T	C	C	0.92	T	T	Halar_0998	0	287	0	25	312	ct	Halar_0998	2507078903	gi|345004282|ref|YP_004807135.1|	arCOG06322	3	E	COG0506	E	Proline dehydrogenase
Contig115	285271	C	T	T	0.964	T	F	Halar_1002	0	9	0	238	247	ac	Halar_1002	2507078907	gi|345004286|ref|YP_004807139.1|	arCOG00349	3	C	COG1141	C	Ferredoxin
Contig115	286364	A	G	G	0.903	T	T	Halar_1003	19	0	176	0	195	aa	Halar_1003	2507078908	gi|345004287|ref|YP_004807140.1|	arCOG00872	1	L	COG1111	L	ERCC4-like helicase
Contig115	286477	C	T	T	0.911	T	F	Halar_1003	0	14	0	143	157	gc	Halar_1003	2507078908	gi|345004287|ref|YP_004807140.1|	arCOG00872	1	L	COG1111	L	ERCC4-like helicase
Contig115	286511	A	G	G	0.974	T	T	Halar_1003	4	0	150	0	154	ca	Halar_1003	2507078908	gi|345004287|ref|YP_004807140.1|	arCOG00872	1	L	COG1111	L	ERCC4-like helicase
Contig115	286523	T	C	C	0.954	T	T	Halar_1003	0	165	0	8	173	tt	Halar_1003	2507078908	gi|345004287|ref|YP_004807140.1|	arCOG00872	1	L	COG1111	L	ERCC4-like helicase
Contig115	286682	C	A	A	0.941	T	T	Halar_1003	239	15	0	0	254	cc	Halar_1003	2507078908	gi|345004287|ref|YP_004807140.1|	arCOG00872	1	L	COG1111	L	ERCC4-like helicase
Contig115	349990	A	G	G	0.952	T	T	Halar_1071	13	0	256	0	269	ca	Halar_1071	2507078977	gi|345004349|ref|YP_004807202.1|	arCOG01301	2	O	COG0492	O	Thioredoxin reductase
Contig115	351625	T	G	G	0.982	T	F	Halar_1074	0	0	222	4	226	at	Halar_1074	2507078980	gi|345004352|ref|YP_004807205.1|	arCOG03873	4	S	COG2855	S	Predicted membrane protein
Contig115	383034	C	T	T	0.988	T	F	Halar_1102	0	4	0	337	341	cc	Halar_1102	2507079008	gi|345004380|ref|YP_004807233.1|	arCOG00373	1	L	COG1196	D	DNA sulfur modification protein DndD, ATPase
Contig115	474990	A	G	G	1	T	F	Halar_1196	0	0	320	0	320	aa									
Contig115	487055	G	T	T	0.913	T	F	Halar_1209	0	0	17	179	196	ag	Halar_1209	2507079115	gi|345004481|ref|YP_004807334.1|	arCOG02092	3	E	COG0119	E	Isopropylmalate/homocitrate /citramalate synthase
Contig115	560597	A	G	G	0.912	T	T	Halar_1281	22	0	227	0	249	ca	Halar_1281	2507079187	gi|345004548|ref|YP_004807401.1|	arCOG00184	3	E	COG4608	E	ABC-type oligopeptide transport system, ATPase component
Contig115	560728	C	T	T	0.935	T	F	Halar_1281	0	16	0	231	247	gc	Halar_1281	2507079187	gi|345004548|ref|YP_004807401.1|	arCOG00184	3	E	COG4608	E	ABC-type oligopeptide transport system, ATPase component
Contig115	560744	G	A	A	0.939	T	T	Halar_1281	229	0	15	0	244	gg	Halar_1281	2507079187	gi|345004548|ref|YP_004807401.1|	arCOG00184	3	E	COG4608	E	ABC-type oligopeptide transport system, ATPase component
Contig115	667430	G	A	A	0.933	T	F	Halar_1394	277	0	20	0	297	gg	Halar_1394	2507079301	gi|345004658|ref|YP_004807511.1|	arCOG01808	2	N	COG2064	NU	Flp pilus assembly protein TadC
Contig115	746010	A	G	G	0.98	F	F		5	0	246	0	251	ca									
Contig115	821561	T	G	G	0.939	T	T	Halar_1558	0	0	307	20	327	ct	Halar_1558	2507079467	gi|345004810|ref|YP_004807663.1|	arCOG01722	1	J	COG0099	J	Ribosomal protein S13
Contig115	851612	A	G	G	0.906	T	F	Halar_1593	28	0	271	0	299	ga	Halar_1593	2507079504	gi|345004844|ref|YP_004807697.1|	arCOG01331	2	O	COG0501	O	Zn-dependent protease with chaperone function
Contig115	903759	C	T	T	0.923	T	T	Halar_1644	0	27	0	322	349	ac	Halar_1644	2507079555	gi|345004889|ref|YP_004807742.1|	arCOG08231	4	S			Uncharacterized conserved protein
Contig115	915519	T	C	C	0.987	T	F	Halar_1654	1	311	0	3	315	gt	Halar_1654	2507079566	gi|345004899|ref|YP_004807752.1|	arCOG10342	4	R			Ferritin-like superfamily protein
Contig115	1029727	G	A	A	0.917	T	T	Halar_1764	165	0	15	0	180	cg	Halar_1764	2507079682	gi|345005001|ref|YP_004807854.1|	arCOG00134	3	G	COG0477	GEPR	Permease of the major facilitator superfamily
Contig115	1030031	A	G	G	0.974	T	F	Halar_1764	6	0	229	0	235	ca	Halar_1764	2507079682	gi|345005001|ref|YP_004807854.1|	arCOG00134	3	G	COG0477	GEPR	Permease of the major facilitator superfamily
Contig115	1030042	G	A	A	0.922	T	T	Halar_1764	200	0	17	0	217	ag	Halar_1764	2507079682	gi|345005001|ref|YP_004807854.1|	arCOG00134	3	G	COG0477	GEPR	Permease of the major facilitator superfamily
Contig115	1030134	C	T	T	0.941	T	F	Halar_1765	0	11	0	176	187	ac	Halar_1765	2507079683	gi|345005002|ref|YP_004807855.1|	arCOG04458	4	R	COG2457	S	Uncharacterized protein of DIM6/NTAB family
Contig115	1030207	C	T	T	0.941	T	F	Halar_1765	0	16	0	254	270	tc	Halar_1765	2507079683	gi|345005002|ref|YP_004807855.1|	arCOG04458	4	R	COG2457	S	Uncharacterized protein of DIM6/NTAB family
Contig115	1030376	T	C	C	0.93	T	T	Halar_1765	1	199	0	14	214	ct	Halar_1765	2507079683	gi|345005002|ref|YP_004807855.1|	arCOG04458	4	R	COG2457	S	Uncharacterized protein of DIM6/NTAB family
Contig115	1030385	A	G	G	0.934	T	T	Halar_1765	15	0	211	0	226	ga	Halar_1765	2507079683	gi|345005002|ref|YP_004807855.1|	arCOG04458	4	R	COG2457	S	Uncharacterized protein of DIM6/NTAB family
Contig115	1090205	C	T	T	0.973	T	F	Halar_1826	0	7	0	249	256	gc	Halar_1826	2507079744	gi|345005056|ref|YP_004807909.1|	arCOG01173	2	T			RecA-superfamily ATPase implicated in signal transduction
Contig115	1102418	A	G	G	1	T	F	Halar_1840	0	0	303	0	303	ca	Halar_1840	2507079759	gi|345005070|ref|YP_004807923.1|	arCOG00998	1	K	COG1958	K	Small nuclear ribonucleoprotein (snRNP) homolog
Contig115	1128855	A	G	G	1	F	F		0	0	179	0	179	ca									
Contig115	1187430	G	C	C	0.908	T	T	Halar_1927	0	148	15	0	163	cg	Halar_1927	2507079848	gi|345005153|ref|YP_004808006.1|	arCOG00777	3	Q	COG2050	Q	HGG motif-containing thioesterase, possibly involved in aromatic compounds catabolism
Contig115	1187610	C	T	T	0.921	T	T	Halar_1927	0	21	0	245	266	gc	Halar_1927	2507079848	gi|345005153|ref|YP_004808006.1|	arCOG00777	3	Q	COG2050	Q	HGG motif-containing thioesterase, possibly involved in aromatic compounds catabolism
Contig115	1188892	T	C	C	0.911	T	T	Halar_1929	0	185	0	18	203	tt	Halar_1929	2507079850	gi|345005155|ref|YP_004808008.1|	arCOG01941	3	H	COG0095	H	Lipoate-protein ligase A
Contig115	1229577	T	G	G	1	T	F	Halar_1978	0	0	238	0	238	tt	Halar_1978	2507079900	gi|345005197|ref|YP_004808050.1|	arCOG06166	1	K			Transcriptional regulator, ArsR family
Contig115	1301302	T	C	C	0.907	T	T	Halar_2054	0	253	0	26	279	tt	Halar_2054	2507079977	gi|345005266|ref|YP_004808119.1|	arCOG02267	3	P	COG0306	P	Phosphate/sulphate permease
Contig115	1320817	C	T	T	0.903	F	F		0	21	0	195	216	gc									
Contig115	1344751	G	A	A	0.911	T	F	Halar_2103	265	0	26	0	291	cg	Halar_2103	2507080027	gi|345005306|ref|YP_004808159.1|	arCOG04302	1	J	COG0008	J	Glutamyl- or glutaminyl-tRNA synthetase
Contig115	1345183	G	A	A	0.948	T	F	Halar_2103	272	0	15	0	287	ag	Halar_2103	2507080027	gi|345005306|ref|YP_004808159.1|	arCOG04302	1	J	COG0008	J	Glutamyl- or glutaminyl-tRNA

																									synthetase
Contig115	1345200	T	C	C	0.983	T	T	Halar_2103	1	282	1	3	287	ct	Halar_2103	2507080027	gi|345005306|ref|YP_004808159.1|	arCOG04302	1	J	COG0008	J	Glutamyl- or glutaminyl-tRNA synthetase
Contig115	1346686	A	G	G	0.946	T	T	Halar_2104	12	0	212	0	224	ca	Halar_2104	2507080028	gi|345005307|ref|YP_004808160.1|	arCOG00915	3	E	COG0160	E	4-aminobutyrate aminotransferase or related aminotransferase
Contig115	1347539	C	T	T	0.916	T	F	Halar_2105	0	25	0	273	298	gc	Halar_2105	2507080029	gi|345005308|ref|YP_004808161.1|	arCOG07989	3	C			Rubredoxin family protein
Contig115	1347648	G	C	C	0.92	F	F		0	276	24	0	300	tg										
Contig115	1395989	T	C	C	0.983	T	T	Halar_2152	0	297	0	5	302	ct	Halar_2152	2507080077	gi|345005353|ref|YP_004808206.1|	arCOG04524	4	S			Uncharacterized conserved protein
Contig115	1428463	A	G	G	0.917	T	T	Halar_2187	26	0	288	0	314	ga	Halar_2187	2507080112	gi|345005388|ref|YP_004808241.1|	arCOG04663	4	S			Uncharacterized conserved protein
Contig115	1462280	G	T	T	1	T	F	Halar_2224	0	0	0	225	225	cg	Halar_2224	2507080149	gi|345005419|ref|YP_004808272.1|	arCOG02202	2	D	COG0206	D	Cell division GTPase
Contig115	1472604	C	T	T	0.942	T	T	Halar_2235	0	11	0	178	189	cc	Halar_2235	2507080160	gi|345005429|ref|YP_004808282.1|	arCOG00102	3	F	COG0047	F	Phosphoribosylformylglycinamidine (FGAM) synthase, glutamine amidotransferase domain
Contig115	1494064	A	G	G	0.903	T	T	Halar_2258	15	0	140	0	155	aa										
Contig115	1549576	A	G	G	1	T	T	Halar_2315	0	0	24	0	24	ta	Halar_2315	2507080241	gi|345005504|ref|YP_004808357.1|	arCOG00235	3	Q	COG0179	Q	2-keto-4-pentenoate hydratase/2-oxohepta-3-ene-1,7-dioic acid hydratase (catechol pathway)
Contig115	1563121	T	C	C	0.906	T	T	Halar_2328	0	259	0	27	286	tt	Halar_2328	2507080254	gi|345005517|ref|YP_004808370.1|	arCOG00271	3	G	COG0697	GER	Permease of the drug/metabolite transporter (DMT) superfamily
Contig115	1563157	C	T	T	0.903	T	T	Halar_2328	0	26	0	242	268	tc	Halar_2328	2507080254	gi|345005517|ref|YP_004808370.1|	arCOG00271	3	G	COG0697	GER	Permease of the drug/metabolite transporter (DMT) superfamily
Contig115	1599568	G	A	A	0.98	T	F	Halar_2359	193	0	4	0	197	ag	Halar_2359	2507080287	gi|345005546|ref|YP_004808399.1|	arCOG00014	3	G	COG0524	G	Sugar kinase, ribokinase family
Contig115	1694407	T	C	C	0.948	T	F	Halar_2470	3	275	0	12	290	tt	Halar_2470	2507080398	gi|345005644|ref|YP_004808497.1|	arCOG04094	1	J	COG0198	J	Ribosomal protein L24
Contig115	1758515	A	C	C	0.954	T	F	Halar_2551	13	271	0	0	284	ta	Halar_2551	2507080480	gi|345005724|ref|YP_004808577.1|	arCOG04817	2	M	COG1083	M	CMP-N-acetylneuraminic acid synthetase
Contig115	1780103	A	G	G	0.922	F	F		22	0	260	0	282	ga										
Contig115	1780146	C	T	T	0.911	F	F		0	24	0	247	271	cc										
Contig115	1780171	C	T	T	0.918	F	F		0	23	0	256	279	cc										
Contig115	1820977	A	G	G	0.977	F	F		6	0	255	0	261	ta										
Contig115	1925417	C	G	G	0.996	F	F		0	1	258	0	259	tc										
Contig115	2033878	T	C	C	0.924	F	F		0	342	0	28	370	at										
Contig115	2048929	G	A	A	0.914	T	T	Halar_2839	235	0	22	0	257	cg	Halar_2839	2507080773	gi|345005991|ref|YP_004808844.1|	arCOG00412	1	J	COG0072	J	Phenylalanyl-tRNA synthetase beta subunit
Contig115	2049244	A	G	G	0.91	T	T	Halar_2839	23	0	233	0	256	ca	Halar_2839	2507080773	gi|345005991|ref|YP_004808844.1|	arCOG00412	1	J	COG0072	J	Phenylalanyl-tRNA synthetase beta subunit
Contig115	2049329	C	G	G	0.921	T	F	Halar_2839	0	22	258	0	280	cc	Halar_2839	2507080773	gi|345005991|ref|YP_004808844.1|	arCOG00412	1	J	COG0072	J	Phenylalanyl-tRNA synthetase beta subunit
Contig115	2095210	T	C	C	0.991	T	T	Halar_2887	1	224	0	1	226	ct	Halar_2887	2507080821	gi|345006038|ref|YP_004808891.1|	arCOG01904	3	H	COG1339	KH	CTP-dependent Riboflavin kinase
Contig115	2096384	A	G	G	0.956	T	T	Halar_2888	12	0	261	0	273	ga	Halar_2888	2507080822	gi|345006039|ref|YP_004808892.1|	arCOG01320	3	H	COG0108	H	3,4-dihydroxy-2-butanone 4-phosphate synthase
Contig115	2105852	G	A	A	0.985	T	T	Halar_2898	133	0	2	0	135	cg	Halar_2898	2507080832	gi|345006049|ref|YP_004808902.1|	arCOG01981	1	K	COG1405	K	Transcription initiation factor TFIIIB, Brf1 subunit/Transcription initiation factor TFIIB
Contig115	2105876	A	G	G	0.984	T	T	Halar_2898	2	0	120	0	122	ca	Halar_2898	2507080832	gi|345006049|ref|YP_004808902.1|	arCOG01981	1	K	COG1405	K	Transcription initiation factor TFIIIB, Brf1 subunit/Transcription initiation factor TFIIB
Contig115	2105894	C	G	G	0.974	T	T	Halar_2898	0	4	148	0	152	tc	Halar_2898	2507080832	gi|345006049|ref|YP_004808902.1|	arCOG01981	1	K	COG1405	K	Transcription initiation factor TFIIIB, Brf1 subunit/Transcription initiation factor TFIIB
Contig115	2106485	G	A	A	0.933	F	F		28	0	2	0	30	gg										
Contig115	2116807	A	G	G	0.992	F	F		2	0	241	0	243	ca										
Contig115	2167873	T	C	C	0.996	T	T	Halar_2957	0	250	0	1	251	tt	Halar_2957	2507080896	gi|345006106|ref|YP_004808959.1|	arCOG04128	3	F	COG0035	F	Uracil phosphoribosyltransferase
Contig115	2178771	C	A	A	0.919	T	T	Halar_2968	238	21	0	0	259	cc	Halar_2968	2507080907	gi|345006117|ref|YP_004808970.1|	arCOG06550	4	S			Uncharacterized conserved protein
Contig115	2219325	C	T	T	0.939	F	F		0	16	0	247	263	ac										
Contig115	2251531	G	A	A	0.902	T	T	Halar_3041	220	0	24	0	244	cg										
Contig115	2262320	T	C	C	0.985	T	F	Halar_3054	0	261	0	4	265	tt	Halar_3054	2507080994	gi|345006199|ref|YP_004809052.1|	arCOG04723	4	S			Uncharacterized conserved protein
Contig115	2272020	A	G	G	0.979	F	F		5	0	230	0	235	ga										
Contig115	2275652	A	G	G	1	T	T	Halar_3068	0	0	293	0	293	ta	Halar_3068	2507081008	gi|345006213|ref|YP_004809066.1|	arCOG01115	1	J	COG1571	R	tRNA(Ile2) 2-agmatinylcytidine synthetase;

																									containing Zn-ribbon domain and OB-fold domain
Contig115	2286191	A	G	G	0.986	T	T	Halar_3078	3	0	280	1	284	ca	Halar_3078	2507081018	gi|345006223|ref|YP_004809076.1|	arCOG04351	4	S	COG3356	S	Predicted membrane protein
Contig115	2313706	T	C	C	0.974	T	F	Halar_3108	0	301	0	8	309	ct	Halar_3108	2507081049	gi|345006249|ref|YP_004809102.1|	arCOG04623	4	R	COG3380	R	Predicted NAD/FAD-dependent oxidoreductase
Contig115	2403615	C	T	T	0.902	F	F		0	16	0	147	163	cc											
Contig115	2419491	A	G	G	0.906	T	F	Halar_3211	21	0	202	0	223	ca	Halar_3211	2507081153	gi|345006345|ref|YP_004809198.1|	arCOG00219	3	P	COG0725	P	ABC-type molybdate transport system, periplasmic component
Contig115	2426544	C	T	T	0.968	T	T	Halar_3217	0	9	0	269	278	gc	Halar_3217	2507081159	gi|345006351|ref|YP_004809204.1|	arCOG01646	3	E			Dipeptidyl aminopeptidase/acylaminoacyl-peptidase
Contig115	2426744	T	C	C	0.944	T	F	Halar_3217	0	238	1	13	252	ct	Halar_3217	2507081159	gi|345006351|ref|YP_004809204.1|	arCOG01646	3	E			Dipeptidyl aminopeptidase/acylaminoacyl-peptidase
Contig115	2453957	G	A	A	0.928	T	F	Halar_3242	258	0	20	0	278	gg	Halar_3242	2507081184	gi|345006375|ref|YP_004809228.1|	arCOG02327	2	T	COG0642	T	Signal transduction histidine kinase, contains PAS domain
Contig115	2455355	T	C	C	0.935	T	F	Halar_3244	0	172	0	12	184	ct	Halar_3244	2507081186	gi|345006377|ref|YP_004809230.1|	arCOG01641	4	R			Predicted RNA-binding protein, contains TRAM domain
Contig115	2455395	G	A	A	0.923	T	T	Halar_3244	169	0	14	0	183	cg	Halar_3244	2507081186	gi|345006377|ref|YP_004809230.1|	arCOG01641	4	R			Predicted RNA-binding protein, contains TRAM domain
Contig115	2543197	T	C	C	0.994	T	F	Halar_3343	0	175	0	1	176	at	Halar_3343	2507081286	gi|345006470|ref|YP_004809323.1|	arCOG10816	4	S	COG5476	S	Uncharacterized conserved protein
Contig115	2553021	C	T	T	0.931	F	F		0	26	0	352	378	cc											
Contig115	2584894	A	G	G	0.951	T	F	Halar_3384	12	0	233	0	245	ga	Halar_3384	2507081330	gi|345006509|ref|YP_004809362.1|	arCOG02998	4	S	COG1917	S	Uncharacterized conserved protein, contains double-stranded beta-helix domain
Contig115	2592586	C	G	G	0.996	T	F	Halar_3393	1	0	259	0	260	gc	Halar_3393	2507081339	gi|345006516|ref|YP_004809369.1|	arCOG02333	2	T			Signal transduction histidine kinase, contains REC and PAS domains
Contig115	2638830	A	C	C	1	T	T	Halar_3436	0	266	0	0	266	ca	Halar_3436	2507081382	gi|345006554|ref|YP_004809407.1|	arCOG01534	3	E			ABC-type transport system, periplasmic component
Contig115	2701888	A	G	G	0.945	T	T	Halar_3506	11	1	205	0	217	ca	Halar_3506	2507081456	gi|345006620|ref|YP_004809473.1|	arCOG01622	4	R			Predicted dehydrogenase
Contig115	2822442	C	T	T	0.953	T	F	Halar_3644	0	13	0	263	276	tc	Halar_3644	2507081595	gi|345006743|ref|YP_004809596.1|	arCOG02847	2	O	COG2214	O	DnaJ-class molecular chaperone
Contig115	2858828	C	G	G	0.912	T	T	Halar_3682	0	20	208	0	228	ac	Halar_3682	2507081633	gi|345006777|ref|YP_004809630.1|	arCOG04799	4	S			Uncharacterized conserved protein
Contig115	2882202	T	C	C	0.927	F	F		0	268	0	21	289	ct											
Contig32	27	T	C	C	0.996	F	F		0	245	1	0	246	ct											
Contig32	130042	A	G	G	1	T	T	halTADL_0138	0	0	354	0	354	ga	halTADL_0138	2507074525	gi|336252565|ref|YP_004595672.1|	arCOG02369	2	T			Signal transduction histidine kinase
Contig32	147274	C	G	G	0.972	F	F		0	9	310	0	319	gc											
Contig32	197582	T	C	C	1	T	F	halTADL_0206	0	179	0	0	179	ct											
Contig32	198183	G	A	A	1	T	T	halTADL_0206	210	0	0	0	210	gg											
Contig32	198430	T	C	C	1	T	F	halTADL_0206	0	201	0	0	201	gt											
Contig32	201265	T	C	C	1	T	F	halTADL_0210	0	42	0	0	42	tt											
Contig32	204178	T	C	C	0.955	F	F		0	21	0	1	22	at											
Contig32	221598	A	C	C	0.9	T	F	halTADL_0233	52	470	0	0	522	ca	halTADL_0233	2507074620	gi|336252969|ref|YP_004596076.1|	arCOG03286	3	G	COG3387	G	Glycosyl hydrolase family 15
Contig32	267903	G	C	C	0.962	T	F	halTADL_0283	0	641	25	0	666	cg	halTADL_0283	2507074670	gi|336253800|ref|YP_004596907.1|	arCOG03256	4	S			Uncharacterized conserved secreted protein, contains OB-fold domain
Contig32	269802	T	G	G	0.912	T	F	halTADL_0283	0	0	529	51	580	gt	halTADL_0283	2507074670	gi|336253800|ref|YP_004596907.1|	arCOG03256	4	S			Uncharacterized conserved secreted protein, contains OB-fold domain
Contig32	333660	T	C	C	0.913	T	F	halTADL_0352	0	462	0	44	506	tt	halTADL_0352	2507074743	gi|76802142|ref|YP_327150.1|	arCOG00005	1	K	COG1695	K	Predicted transcriptional regulator, PadR family
Contig32	346804	T	C	C	0.909	T	F	halTADL_0367	0	472	0	47	519	ct	halTADL_0367	2507074758	gi|284161738|ref|YP_003400361.1|	arCOG02838	2	T			FIST_N domain containing protein
Contig32	379199	C	G	G	1	T	F	halTADL_0393	0	0	492	0	492	gc	halTADL_0393	2507074784	gi|76802448|ref|YP_327456.1|	arCOG06337	2	M	COG4948	MR	L-alanine-DL-glutamate epimerase or related enzyme of enolase superfamily
Contig32	421686	T	G	G	0.998	T	T	halTADL_0433	0	0	468	1	469	gt	halTADL_0433	2507074824	gi|336252121|ref|YP_004595228.1|	arCOG02327	2	T	COG0642	T	Signal transduction histidine kinase, contains PAS domain
Contig32	443214	T	C	C	0.998	T	F	halTADL_0453	1	469	0	0	470	gt											
Contig32	468733	A	G	G	1	T	F	halTADL_0482	0	0	631	0	631	ta	halTADL_0482	2507074874	gi|284164660|ref|YP_003402939.1|	arCOG06227	4	S			Uncharacterized conserved protein
Contig32	726848	A	G	G	0.997	F	F		2	0	598	0	600	ta											
Contig32	728342	T	C	C	0.94	F	F		0	375	0	24	399	tt											
Contig32	729183	A	G	G	0.933	F	F		33	0	462	0	495	ta											
Contig32	729274	T	A	A	0.915	F	F		476	0	1	43	520	ct											
Contig32	762917	A	G	G	1	F	F		0	0	462	0	462	ca											

Contig32	763749	C	A	A	0.913	T	T	halTADL_0783	272	26	0	0	298	gc	halTADL_0783	2507075183	gi|284164212|ref|YP_003402491.1|	arCOG00571	3	C	COG1053	C	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit
Contig32	763752	G	C	C	0.91	T	T	halTADL_0783	0	274	27	0	301	cg	halTADL_0783	2507075183	gi|284164212|ref|YP_003402491.1|	arCOG00571	3	C	COG1053	C	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit
Contig32	763911	T	C	C	0.904	T	T	halTADL_0783	0	283	0	30	313	ct	halTADL_0783	2507075183	gi|284164212|ref|YP_003402491.1|	arCOG00571	3	C	COG1053	C	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit
Contig32	764148	C	T	T	0.917	T	T	halTADL_0783	0	42	0	467	509	cc	halTADL_0783	2507075183	gi|284164212|ref|YP_003402491.1|	arCOG00571	3	C	COG1053	C	Succinate dehydrogenase/fumarate reductase, flavoprotein subunit
Contig32	772370	T	C	C	0.948	F	F		0	589	0	32	621	tt										
Contig32	772868	T	C	C	0.933	T	T	halTADL_0792	0	491	0	35	526	gt	halTADL_0792	2507075192	gi|344211538|ref|YP_004795858.1|	arCOG01112	3	G			PEP-utilising enzyme associated, "swivelling" beta/beta/alpha domain
Contig32	774851	T	C	C	0.913	F	F		0	547	0	52	599	gt										
Contig32	774892	A	G	G	0.928	T	T	halTADL_0794	38	0	492	0	530	ca	halTADL_0794	2507075194	gi|284172618|ref|YP_003406000.1|	arCOG06426	4	S			Uncharacterized conserved protein
Contig32	790320	T	A	A	0.927	T	F	halTADL_0813	536	0	0	42	578	ct	halTADL_0813	2507075213	gi|284167228|ref|YP_003405506.1|	arCOG01619	4	R	COG0656	R	Aldo/keto reductase, related to diketogulonate reductase
Contig32	795594	G	A	A	0.909	T	T	halTADL_0819	401	0	40	0	441	cg	halTADL_0819	2507075219	gi|222480484|ref|YP_002566721.1|	arCOG00403	1	J	COG0172	J	Seryl-tRNA synthetase
Contig32	796047	G	A	A	0.908	T	T	halTADL_0819	445	0	45	0	490	cg	halTADL_0819	2507075219	gi|222480484|ref|YP_002566721.1|	arCOG00403	1	J	COG0172	J	Seryl-tRNA synthetase
Contig32	798775	A	G	G	0.913	F	F		40	0	422	0	462	ca										
Contig32	809277	A	G	G	1	F	F		0	0	411	0	411	ta										
Contig32	810099	T	C	C	0.906	T	F	halTADL_0831	0	444	0	46	490	tt										
Contig32	824105	T	G	G	0.994	T	F	halTADL_0848	1	0	462	2	465	gt	halTADL_0848	2507075248	gi|55379917|ref|YP_137766.1|	arCOG01007	3	P	COG0609	P	ABC-type Fe3+-siderophore transport system, permease component
Contig32	858302	G	C	C	0.996	T	T	halTADL_0886	0	541	2	0	543	ag	halTADL_0886	2507075286	gi|336252493|ref|YP_004595600.1|	arCOG00086	3	E	COG0512	EH	Anthranilate/para-aminobenzoate synthase component II
Contig32	923008	C	G	G	0.998	F	F		0	1	534	0	535	ac										
Contig32	997271	T	C	C	0.991	T	T	halTADL_1034	1	448	0	3	452	ct	halTADL_1034	2507075437	gi|292655074|ref|YP_003534971.1|	arCOG00358	4	R	COG1163	R	Predicted GTPase
Contig32	1003922	C	G	G	0.957	T	F	halTADL_1039	0	8	179	0	187	ac	halTADL_1039	2507075442	gi|76801877|ref|YP_326885.1|	arCOG06273	4	S			S-layer protein
Contig32	1012659	G	C	C	0.942	T	T	halTADL_1044	0	65	4	0	69	tg	halTADL_1044	2507075447	gi|313126785|ref|YP_004037055.1|	arCOG02053	2	T	COG0589	T	Nucleotide-binding protein, UspA family
Contig32	1012773	T	C	C	0.947	T	T	halTADL_1044	0	162	0	9	171	at	halTADL_1044	2507075447	gi|313126785|ref|YP_004037055.1|	arCOG02053	2	T	COG0589	T	Nucleotide-binding protein, UspA family
Contig32	1013400	A	G	G	0.966	T	T	halTADL_1045	5	0	140	0	145	ta	halTADL_1045	2507075448	gi|222479329|ref|YP_002565566.1|	arCOG04471	2	U	COG4083	S	Exosortase
Contig32	1013607	T	C	C	0.946	T	T	halTADL_1045	0	210	0	12	222	at	halTADL_1045	2507075448	gi|222479329|ref|YP_002565566.1|	arCOG04471	2	U	COG4083	S	Exosortase
Contig32	1013835	A	C	C	0.961	T	T	halTADL_1045	8	198	0	0	206	ca	halTADL_1045	2507075448	gi|222479329|ref|YP_002565566.1|	arCOG04471	2	U	COG4083	S	Exosortase
Contig32	1014174	T	G	G	0.951	T	T	halTADL_1046	0	0	39	2	41	ct	halTADL_1046	2507075449	gi|292655079|ref|YP_003534976.1|	arCOG04161	1	J	COG1798	J	Diphthamide biosynthesis methyltransferase
Contig32	1014354	A	G	G	0.902	T	T	halTADL_1046	6	0	55	0	61	ca	halTADL_1046	2507075449	gi|292655079|ref|YP_003534976.1|	arCOG04161	1	J	COG1798	J	Diphthamide biosynthesis methyltransferase
Contig32	1014745	A	G	G	0.929	T	F	halTADL_1046	6	0	79	0	85	ca	halTADL_1046	2507075449	gi|292655079|ref|YP_003534976.1|	arCOG04161	1	J	COG1798	J	Diphthamide biosynthesis methyltransferase
Contig32	1015122	T	C	C	0.942	T	T	halTADL_1047	0	81	0	5	86	gt	halTADL_1047	2507075450	gi|222479340|ref|YP_002565577.1|	arCOG06262	4	S			Uncharacterized conserved protein
Contig32	1015233	G	C	C	0.949	T	T	halTADL_1047	0	75	4	0	79	tg	halTADL_1047	2507075450	gi|222479340|ref|YP_002565577.1|	arCOG06262	4	S			Uncharacterized conserved protein
Contig32	1015620	G	C	C	0.912	T	T	halTADL_1047	0	31	3	0	34	cg	halTADL_1047	2507075450	gi|222479340|ref|YP_002565577.1|	arCOG06262	4	S			Uncharacterized conserved protein
Contig32	1016009	A	C	C	0.913	T	T	halTADL_1048	2	21	0	0	23	ca	halTADL_1048	2507075451	gi|313127071|ref|YP_004037341.1|	arCOG00570	3	C			Dehydrogenase (flavoprotein
Contig32	1101056	T	G	G	0.989	T	F	halTADL_1135	0	0	459	5	464	at	halTADL_1135	2507075540	gi|345005893|ref|YP_004808746.1|	arCOG03104	4	S			Uncharacterized conserved protein
Contig32	1116980	T	C	C	1	T	F	halTADL_1150	0	381	0	0	381	at	halTADL_1150	2507075555	gi|345005859|ref|YP_004808712.1|	arCOG00609	2	M	COG0750	M	Predicted membrane-associated Zn-dependent protease
Contig32	1124845	A	G	G	0.993	T	T	halTADL_1162	3	0	409	0	412	ta										
Contig32	1124854	G	A	A	0.983	T	T	halTADL_1162	393	0	6	1	400	cg										
Contig32	1184589	A	G	G	0.995	T	F	halTADL_1220	0	0	195	1	196	ca	halTADL_1220	2507075626	gi|292655843|ref|YP_003535740.1|	arCOG04386	3	C	COG1620	C	L-lactate permease
Contig32	1186013	C	G	G	1	T	F	halTADL_1222	0	0	224	0	224	tc										
Contig32	1205576	T	C	C	0.997	T	F	halTADL_1247	1	353	0	0	354	gt	halTADL_1247	2507075653	gi|300711034|ref|YP_003736848.1|	arCOG00467	1	L	COG1474	LO	Cdc6-related protein, AAA superfamily ATPase
Contig32	1228580	C	G	G	1	T	F	halTADL_1268	0	0	68	0	68	gc	halTADL_1268	2507075674	gi|76803390|ref|YP_327659.1|	arCOG06448	4	S			Uncharacterized conserved protein
Contig32	1252636	G	A	A	1	F	F		37	0	0	0	37	ag										

Contig32	1299011	A	G	G	0.998	T	F	halTADL_1348	1	0	614	0	615	ta	halTADL_1348	2507075754	gi|257053041|ref|YP_003130874.1|	arCOG04064	2	M			Predicted membrane-associated Zn-dependent protease
Contig32	1345405	A	G	G	1	F	F		0	0	99	0	99	ta											
Contig32	1374401	A	G	G	0.912	T	F	halTADL_1422	13	0	134	0	147	ca	halTADL_1422	2507075828	gi|345006615|ref|YP_004809468.1|	arCOG03560	4	S			Uncharacterized conserved protein
Contig32	1467380	G	A	A	0.901	T	T	halTADL_1514	154	0	17	0	171	gg	halTADL_1514	2507075920	gi|313117376|ref|YP_004044359.1|	arCOG05365	3	I			Oligosaccharyl transferase STT3 subunit related protein
Contig32	1532733	G	A	A	0.902	T	T	halTADL_1590	632	1	68	0	701	ag	halTADL_1590	2507075996	gi|345133542|ref|YP_004821333.1|	arCOG07786	1	K			Predicted Helicase fused to HTH domain
Contig32	1572162	A	G	G	0.998	T	F	halTADL_1627	1	0	564	0	565	ca	halTADL_1627	2507076034	gi|257386351|ref|YP_003176124.1|	arCOG02320	2	N			Methyl-accepting chemotaxis protein
Contig32	1580276	T	C	C	0.998	T	T	halTADL_1636	0	521	0	1	522	tt	halTADL_1636	2507076043	gi|292654994|ref|YP_003534891.1|	arCOG01226	3	E	COG1703	E	Putative periplasmic protein kinase ArgK or related GTPase of G3E family
Contig32	1581133	A	G	G	1	T	T	halTADL_1638	0	0	517	0	517	aa	halTADL_1638	2507076045	gi|289580799|ref|YP_003479265.1|	arCOG01134	3	E	COG0436	E	Aspartate/tyrosine/aromatic aminotransferase
Contig32	1593404	T	C	C	1	T	F	halTADL_1654	0	518	0	0	518	at	halTADL_1654	2507076061	gi|292654901|ref|YP_003534798.1|	arCOG06212	4	S			Uncharacterized conserved protein
Contig32	1621702	A	G	G	0.996	T	F	halTADL_1680	1	0	227	0	228	ta	halTADL_1680	2507076087	gi|292494093|ref|YP_003533236.1|	arCOG03902	1	L			Transposase
Contig32	1649558	C	T	T	0.998	F	F		1	0	0	512	513	cc											
Contig32	1654847	C	G	G	0.996	F	F		0	2	452	0	454	cc											
Contig32	1674277	T	C	C	0.998	T	T	halTADL_1742	0	434	1	0	435	tt	halTADL_1742	2507076150	gi|55379271|ref|YP_137121.1|	arCOG00067	3	F	COG0462	FE	Phosphoribosylpyrophosphate synthetase
Contig32	1678908	T	C	C	1	T	F	halTADL_1748	0	429	0	0	429	ct	halTADL_1748	2507076156	gi|313124906|ref|YP_004035170.1|	arCOG00041	4	R	COG1926	R	Predicted phosphoribosyltransferase
Contig32	1769210	A	G	G	0.995	T	F	halTADL_1838	2	1	575	0	578	ta	halTADL_1838	2507076246	gi|222480670|ref|YP_002566907.1|	arCOG02395	2	N	COG0835	NT	Chemotaxis signal transduction protein
Contig32	1769934	T	C	C	1	F	F		0	603	0	0	603	ct											
Contig32	1770705	A	G	G	1	F	F		0	0	406	0	406	ca											
Contig32	1939040	T	C	C	1	T	F	halTADL_2014	0	98	0	0	98	gt	halTADL_2014	2507076424	gi|336251627|ref|YP_004598858.1|	arCOG09003	4	S			Uncharacterized conserved protein
Contig32	2048957	A	G	G	0.981	T	T	halTADL_2141	9	0	472	0	481	ca	halTADL_2141	2507076551	gi|257052607|ref|YP_003130440.1|	arCOG00266	4	R			Sulfite oxidase or related enzyme
Contig32	2049131	C	T	T	0.99	T	T	halTADL_2141	0	5	0	515	520	tc	halTADL_2141	2507076551	gi|257052607|ref|YP_003130440.1|	arCOG00266	4	R			Sulfite oxidase or related enzyme
Contig32	2058128	T	C	C	0.96	T	T	halTADL_2151	0	499	0	21	520	ct	halTADL_2151	2507076561	gi|55378844|ref|YP_136694.1|	arCOG00318	3	P	COG0704	P	Phosphate uptake regulator
Contig32	2062298	T	C	C	0.944	T	F	halTADL_2155	1	286	0	16	303	tt	halTADL_2155	2507076565	gi|313125578|ref|YP_004035842.1|	arCOG00213	3	P	COG0226	P	ABC-type phosphate transport system, periplasmic component
Contig32	2063372	A	G	G	0.947	F	F		29	0	517	0	546	ca											
Contig32	2076214	T	C	C	0.998	T	T	halTADL_2168	0	456	0	1	457	ct	halTADL_2168	2507076578	gi|222480481|ref|YP_002566718.1|	arCOG01701	1	J			tRNA splicing endonuclease
Contig32	2165082	T	C	C	1	T	T	halTADL_2249	0	485	0	0	485	at	halTADL_2249	2507076666	gi|110667688|ref|YP_657499.1|	arCOG00024	3	C	COG0554	C	Glycerol kinase
Contig32	2165893	A	C	C	0.945	T	F	halTADL_2249	26	445	0	0	471	ga	halTADL_2249	2507076666	gi|110667688|ref|YP_657499.1|	arCOG00024	3	C	COG0554	C	Glycerol kinase
Contig32	2166321	A	G	G	0.988	T	F	halTADL_2250	5	0	505	1	511	ga	halTADL_2250	2507076667	gi|300711496|ref|YP_003737310.1|	arCOG04584	4	S			Uncharacterized conserved protein
Contig32	2188888	A	G	G	1	T	F	halTADL_2275	0	0	130	0	130	ca	halTADL_2275	2507076692	gi|345005189|ref|YP_004808042.1|	arCOG01743	1	J	COG1503	J	Peptide chain release factor 1 (eRF1)
Contig32	2208047	A	G	G	0.908	T	F	halTADL_2296	49	0	495	1	545	ga	halTADL_2296	2507076713	gi|292656745|ref|YP_003536642.1|	arCOG04783	3	P	COG0569	P	TrkA, K+ transport system, NAD-binding component
Contig32	2223689	A	G	G	0.969	T	T	halTADL_2309	13	1	462	1	477	ga	halTADL_2309	2507076726	gi|313127513|ref|YP_004037783.1|	arCOG00757	3	E	COG0404	E	Glycine cleavage system T protein (aminomethyltransferase)
Contig32	2263765	G	C	C	0.933	T	T	halTADL_2352	0	56	4	0	60	gg	halTADL_2352	2507076770	gi|76801912|ref|YP_326920.1|	arCOG02814	3	Q	COG2124	Q	Cytochrome P450
Contig32	2295886	T	G	G	1	F	F		0	0	77	0	77	at											
Contig32	2295897	T	G	G	0.909	F	F		0	0	70	7	77	tt											
Contig32	2296794	C	G	G	1	T	F	halTADL_2389	0	0	197	0	197	cc											
Contig32	2320353	T	C	C	0.923	T	F	halTADL_2408	0	36	0	3	39	at	halTADL_2408	2507076826	gi|344210289|ref|YP_004786465.1|	arCOG01403	2	M			Glycosyltransferase
Contig32	2328402	A	G	G	1	T	F	halTADL_2415	0	0	264	0	264	ta	halTADL_2415	2507076833	gi|344211125|ref|YP_004795445.1|	arCOG00546	1	J	COG0595	R	mRNA degradation ribonuclease J1/J2 (metallo-beta-lactamase superfamily)
Contig32	2332265	T	A	A	0.909	T	F	halTADL_2418	579	0	0	58	637	gt	halTADL_2418	2507076836	gi|222480519|ref|YP_002566756.1|	arCOG00982	3	C	COG0371	C	Glycerol dehydrogenase or related enzyme
Contig32	2350562	T	C	C	0.917	T	T	halTADL_2434	0	319	0	29	348	tt	halTADL_2434	2507076852	gi|336254250|ref|YP_004597357.1|	arCOG01308	2	O			ATPase of the AAA+ class, CDC48 family
Contig32	2351003	A	G	G	0.914	T	T	halTADL_2434	31	0	331	0	362	aa	halTADL_2434	2507076852	gi|336254250|ref|YP_004597357.1|	arCOG01308	2	O			ATPase of the AAA+ class, CDC48 family
Contig32	2351597	C	T	T	0.907	T	T	halTADL_2434	0	31	0	304	335	tc	halTADL_2434	2507076852	gi|336254250|ref|YP_004597357.1|	arCOG01308	2	O			ATPase of the AAA+ class, CDC48 family
Contig32	2351603	G	A	A	0.918	T	T	halTADL_2434	312	0	28	0	340	ag	halTADL_2434	2507076852	gi|336254250|ref|YP_004597357.1|	arCOG01308	2	O			ATPase of the AAA+ class, CDC48 family
Contig32	2352542	T	C	C	0.952	T	F	halTADL_2435	0	315	0	16	331	tt	halTADL_2435	2507076853	gi|222479025|ref|YP_002565262.1|	arCOG00800	1	L	COG1468	L	RecB family exonuclease
Contig32	2352609	T	C	C	0.93	T	T	halTADL_2435	0	387	1	28	416	ct	halTADL_2435	2507076853	gi|222479025|ref|YP_002565262.1|	arCOG00800	1	L	COG1468	L	RecB family exonuclease

Contig32	2353293	A	G	G	0.956	T	T	halTADL_2436	25	0	539	0	564	aa	halTADL_2436	2507076854	gi|345006366|ref|YP_004809219.1|	arCOG04725	1	K			Predicted transcriptional regulator , contains HTH domain
Contig32	2353446	A	G	G	0.941	F	F		31	0	515	1	547	aa											
Contig32	2354488	T	G	G	0.947	T	T	halTADL_2438	0	0	285	16	301	ct	halTADL_2438	2507076856	gi|313124857|ref|YP_004035121.1|	arCOG00202	3	P	COG1122	P	ABC-type cobalt transport system, ATPase component
Contig32	2354989	A	G	G	0.937	T	F	halTADL_2439	28	0	413	0	441	ca	halTADL_2439	2507076857	gi|289580568|ref|YP_003479034.1|	arCOG02250	3	P	COG0619	P	ABC-type cobalt transport system, permease component CbiQ or related transporter
Contig32	2355098	C	T	T	0.948	T	T	halTADL_2439	2	26	0	512	540	cc	halTADL_2439	2507076857	gi|289580568|ref|YP_003479034.1|	arCOG02250	3	P	COG0619	P	ABC-type cobalt transport system, permease component CbiQ or related transporter
Contig32	2355799	C	T	T	0.935	T	F	halTADL_2440	1	27	0	401	429	cc											
Contig32	2355804	A	C	C	0.934	T	T	halTADL_2440	28	399	0	0	427	ca											
Contig32	2355936	T	A	A	0.929	T	F	halTADL_2441	537	0	0	41	578	at											
Contig32	2356249	G	A	A	0.91	T	T	halTADL_2442	488	0	48	0	536	gg											
Contig32	2356381	G	C	C	0.933	T	T	halTADL_2442	0	459	33	0	492	ag											
Contig32	2356573	G	T	T	0.927	T	T	halTADL_2442	0	0	32	404	436	cg											
Contig32	2358443	T	C	C	0.93	T	F	halTADL_2446	0	265	1	19	285	at	halTADL_2446	2507076864	gi|222479156|ref|YP_002565393.1|	arCOG00090	3	F	COG0518	F	GMP synthase - Glutamine amidotransferase domain
Contig32	2376257	C	T	T	1	F	F		0	0	0	102	102	ac											
Contig32	2427080	A	C	C	1	F	F		0	489	0	0	489	ca											
Contig32	2449960	T	C	C	1	T	F	halTADL_2537	0	537	0	0	537	gt											
Contig32	2540449	A	G	G	1	T	T	halTADL_2621	0	0	102	0	102	ca	halTADL_2621	2507077041	gi|55376809|ref|YP_134660.1|	arCOG00260	4	R	COG4603	R	ABC-type uncharacterized transport system, permease component
Contig32	2573920	T	C	C	0.902	F	F		0	450	0	49	499	at											
Contig32	2581081	C	G	G	0.928	F	F		0	43	556	0	599	ac											
Contig32	2582317	C	G	G	0.975	T	T	halTADL_2660	1	11	465	0	477	gc	halTADL_2660	2507077080	gi|292653834|ref|YP_003533732.1|	arCOG00025	3	G	COG1070	G	Sugar (pentulose and hexulose) kinase
Contig32	2586718	T	C	C	0.906	T	T	halTADL_2663	0	416	0	43	459	at	halTADL_2663	2507077083	gi|292653828|ref|YP_003533726.1|	arCOG04585	3	E	COG3075	E	Anaerobic glycerol-3-phosphate dehydrogenase
Contig32	2658867	A	C	C	0.968	F	F		9	276	0	0	285	aa											
Contig32	2663841	C	T	T	0.953	T	T	halTADL_2743	0	19	0	384	403	gc	halTADL_2743	2507077164	gi|289583595|ref|YP_003482005.1|	arCOG00134	3	G	COG0477	GEPR	Permease of the major facilitator superfamily
Contig32	2664796	T	C	C	0.936	T	T	halTADL_2745	0	509	0	35	544	ct											
Contig32	2665690	T	A	A	0.938	F	F		602	0	0	40	642	tt											
Contig32	2665984	T	C	C	0.909	T	T	halTADL_2747	1	231	0	22	254	gt	halTADL_2747	2507077168	gi|313125958|ref|YP_004036228.1|	arCOG04302	1	J	COG0008	J	Glutamyl- or glutaminyl-tRNA synthetase
Contig32	2667175	A	G	G	0.931	T	F	halTADL_2747	31	1	444	1	477	ca	halTADL_2747	2507077168	gi|313125958|ref|YP_004036228.1|	arCOG04302	1	J	COG0008	J	Glutamyl- or glutaminyl-tRNA synthetase
Contig32	2668180	T	C	C	0.926	T	F	halTADL_2748	0	326	0	26	352	ct	halTADL_2748	2507077169	gi|257053426|ref|YP_003131259.1|	arCOG04687	4	R			Ferritin family protein
Contig32	2681008	T	C	C	0.913	T	T	halTADL_2759	1	515	0	48	564	ct	halTADL_2759	2507077180	gi|313125995|ref|YP_004036265.1|	arCOG00500	4	R	COG1234	R	Metal-dependent hydrolase of the beta-lactamase superfamily
Contig32	2725521	G	A	A	1	T	F	halTADL_2809	511	0	0	0	511	gg	halTADL_2809	2507077232	gi|222480123|ref|YP_002566360.1|	arCOG00449	2	T	COG0589	T	Nucleotide-binding protein, UspA family
Contig32	2728323	A	G	G	0.979	F	F		9	0	422	0	431	ca											
Contig32	2738180	C	G	G	0.989	T	F	halTADL_2822	1	0	93	0	94	tc											
Contig32	2750969	T	C	C	0.997	T	F	halTADL_2834	0	345	0	1	346	tt											
Contig32	2776909	A	G	G	1	T	F	halTADL_2866	0	0	277	0	277	ca	halTADL_2866	2507077289	gi|222480713|ref|YP_002566950.1|	arCOG00065	2	O	COG0520	E	Selenocysteine lyase/Cysteine desulfurase
Contig32	2873597	T	C	C	0.902	T	T	halTADL_2970	0	431	1	46	478	tt											
Contig32	2942227	G	C	C	0.998	T	F	halTADL_3040	1	436	0	0	437	gg	halTADL_3040	2507077468	gi|222479707|ref|YP_002565944.1|	arCOG04640	4	S			Uncharacterized conserved protein
Contig32	2950397	A	G	G	1	T	F	halTADL_3051	0	0	483	0	483	ta	halTADL_3051	2507077479	gi|292656319|ref|YP_003536216.1|	arCOG02826	3	F	COG0788	F	Formyltetrahydrofolate hydrolase
Contig32	2983715	G	A	A	0.908	T	F	halTADL_3085	395	0	40	0	435	gg	halTADL_3085	2507077514	gi|289583584|ref|YP_003481994.1|	arCOG00147	3	G	COG0477	GEPR	Permease of the major facilitator superfamily
Contig32	2988787	T	C	C	0.916	F	F		0	439	0	40	479	at											
Contig32	2989631	A	G	G	0.998	T	F	halTADL_3092	0	0	463	1	464	ga	halTADL_3092	2507077521	gi|292655139|ref|YP_003535036.1|	arCOG01554	3	C			F420H2 dehydrogenase subunit, related to NADH:ubiquinone oxidoreductase 20 kD subunit
Contig32	3046200	T	C	C	1	T	F	halTADL_3150	0	504	0	0	504	tt	halTADL_3150	2507077581	gi|222479413|ref|YP_002565650.1|	arCOG08119	4	R	COG3211	R	Secreted phosphatase, PhoX family
Contig32	3132338	A	G	G	0.989	T	F	halTADL_3243	7	0	638	0	645	ca	halTADL_3243	2507077674	gi|345004644|ref|YP_004807497.1|	arCOG04061	1	K	COG1308	K	Transcription factor homologous to NACalpha-BTF3
Contig32	3156757	T	G	G	0.996	F	F		0	0	469	2	471	gt											
Contig32	3168983	G	C	C	0.924	T	T	halTADL_3288	2	477	32	5	516	cg	halTADL_3288	2507077719	gi|284166026|ref|YP_003404305.1|	arCOG10196	4	S	COG1299	G	Uncharacterized conserved protein
Contig32	3178663	A	G	G	0.968	T	T	halTADL_3297	17	0	520	0	537	ca											

HalDL1_Contig37	3828	C	A	A	0.987	T	F	HalDL1_3056	153	2	0	0	155	gc	HalDL1_3056	2507057197	gi|345007177|ref|YP_004810029.1|	arCOG07742	4	S			Uncharacterized conserved protein
HalDL1_Contig37	13562	C	G	G	0.977	T	F	HalDL1_3064	0	4	171	0	175	cc	HalDL1_3064	2507057205	gi|257052527|ref|YP_003130360.1|	arCOG01445	2	V	COG1203	R	CRISPR-associated helicase Cas3
HalDL1_Contig37	20982	G	A	A	0.953	T	T	HalDL1_3068	81	0	4	0	85	ag	HalDL1_3068	2507057209	gi|345007408|ref|YP_004810260.1|	arCOG06564	4	S			Uncharacterized conserved protein
HalDL1_Contig37	20985	T	C	C	0.976	T	T	HalDL1_3068	0	81	0	2	83	tt	HalDL1_3068	2507057209	gi|345007408|ref|YP_004810260.1|	arCOG06564	4	S			Uncharacterized conserved protein
HalDL1_Contig37	128178	C	T	T	0.924	T	T	HalDL1_3179	0	11	0	134	145	ac	HalDL1_3179	2507057320	gi|292653597|ref|YP_003533493.1|	arCOG08895	4	S			Uncharacterized conserved protein
HalDL1_Contig37	168583	A	G	G	0.98	T	T	HalDL1_3225	2	0	98	0	100	ta	HalDL1_3225	2507057366	gi|292656395|ref|YP_003536292.1|	arCOG01764	1	K	COG2101	K	TATA-box binding protein (TBP), component of TFIID and TFIIIB
HalDL1_Contig37	168613	G	A	A	0.996	T	T	HalDL1_3225	222	0	1	0	223	cg	HalDL1_3225	2507057366	gi|292656395|ref|YP_003536292.1|	arCOG01764	1	K	COG2101	K	TATA-box binding protein (TBP), component of TFIID and TFIIIB
HalDL1_Contig37	168661	A	G	G	0.988	T	T	HalDL1_3225	2	0	246	1	249	ca	HalDL1_3225	2507057366	gi|292656395|ref|YP_003536292.1|	arCOG01764	1	K	COG2101	K	TATA-box binding protein (TBP), component of TFIID and TFIIIB
HalDL1_Contig37	168724	T	C	C	0.971	T	T	HalDL1_3225	0	66	0	2	68	at	HalDL1_3225	2507057366	gi|292656395|ref|YP_003536292.1|	arCOG01764	1	K	COG2101	K	TATA-box binding protein (TBP), component of TFIID and TFIIIB
HalDL1_Contig37	168737	C	T	T	0.933	T	F	HalDL1_3225	0	1	0	14	15	gc	HalDL1_3225	2507057366	gi|292656395|ref|YP_003536292.1|	arCOG01764	1	K	COG2101	K	TATA-box binding protein (TBP), component of TFIID and TFIIIB
HalDL1_Contig37	169962	A	G	G	0.966	T	T	HalDL1_3227	2	0	57	0	59	ga	HalDL1_3227	2507057368	gi|298674426|ref|YP_003726176.1|	arCOG00879	2	V	COG0610	V	Type I site-specific restriction-modification system, R (restriction) subunit or related helicase
HalDL1_Contig37	169975	T	C	C	0.987	T	F	HalDL1_3227	0	76	0	1	77	ct	HalDL1_3227	2507057368	gi|298674426|ref|YP_003726176.1|	arCOG00879	2	V	COG0610	V	Type I site-specific restriction-modification system, R (restriction) subunit or related helicase
HalDL1_Contig37	170328	A	T	T	0.966	T	T	HalDL1_3227	4	0	0	112	116	aa	HalDL1_3227	2507057368	gi|298674426|ref|YP_003726176.1|	arCOG00879	2	V	COG0610	V	Type I site-specific restriction-modification system, R (restriction) subunit or related helicase
HalDL1_Contig37	170370	A	G	G	0.973	T	T	HalDL1_3227	5	0	179	0	184	ca	HalDL1_3227	2507057368	gi|298674426|ref|YP_003726176.1|	arCOG00879	2	V	COG0610	V	Type I site-specific restriction-modification system, R (restriction) subunit or related helicase
HalDL1_Contig37	170493	T	C	C	0.977	T	T	HalDL1_3227	1	128	0	2	131	tt	HalDL1_3227	2507057368	gi|298674426|ref|YP_003726176.1|	arCOG00879	2	V	COG0610	V	Type I site-specific restriction-modification system, R (restriction) subunit or related helicase
HalDL1_Contig37	170514	G	A	A	0.988	T	T	HalDL1_3227	169	0	2	0	171	cg	HalDL1_3227	2507057368	gi|298674426|ref|YP_003726176.1|	arCOG00879	2	V	COG0610	V	Type I site-specific restriction-modification system, R (restriction) subunit or related helicase
HalDL1_Contig37	170577	A	G	G	0.984	T	T	HalDL1_3227	3	0	180	0	183	ca	HalDL1_3227	2507057368	gi|298674426|ref|YP_003726176.1|	arCOG00879	2	V	COG0610	V	Type I site-specific restriction-modification system, R (restriction) subunit or related helicase
HalDL1_Contig37	170619	G	C	C	0.937	T	T	HalDL1_3227	0	59	4	0	63	cg	HalDL1_3227	2507057368	gi|298674426|ref|YP_003726176.1|	arCOG00879	2	V	COG0610	V	Type I site-specific restriction-modification system, R (restriction) subunit or related helicase
HalDL1_Contig37	171114	G	R	A	0.907	T	T	HalDL1_3227	39	0	4	0	43	gg	HalDL1_3227	2507057368	gi|298674426|ref|YP_003726176.1|	arCOG00879	2	V	COG0610	V	Type I site-specific restriction-modification system, R (restriction) subunit or related helicase
HalDL1_Contig37	171132	A	G	G	0.975	T	T	HalDL1_3227	2	0	79	0	81	ca	HalDL1_3227	2507057368	gi|298674426|ref|YP_003726176.1|	arCOG00879	2	V	COG0610	V	Type I site-specific restriction-modification system, R (restriction) subunit or related helicase
HalDL1_Contig37	171273	A	G	G	0.984	T	T	HalDL1_3227	1	0	63	0	64	ga	HalDL1_3227	2507057368	gi|298674426|ref|YP_003726176.1|	arCOG00879	2	V	COG0610	V	Type I site-specific restriction-modification system, R (restriction) subunit or related helicase
HalDL1_Contig37	171300	C	T	T	0.955	T	T	HalDL1_3227	0	2	0	42	44	gc	HalDL1_3227	2507057368	gi|298674426|ref|YP_003726176.1|	arCOG00879	2	V	COG0610	V	Type I site-specific restriction-modification system, R (restriction) subunit or related helicase
HalDL1_Contig37	174803	A	G	G	0.965	T	T	HalDL1_3229	5	0	137	0	142	ga	HalDL1_3229	2507057370	gi|298674424|ref|YP_003726174.1|	arCOG05282	2	V			Type I restriction-modification system methyltransferase

																								subunit
HalDL1_Contig37	174821	T	C	C	0.966	T	T	HalDL1_3229	0	169	0	6	175	at	HalDL1_3229	250705737	0gi|298674424|ref|YP_003726174.1|	arCOG05282	2	V			Type I restriction-modification system methyltransferase subunit
HalDL1_Contig37	175622	C	A	A	0.99	T	T	HalDL1_3229	204	2	0	0	206	cc	HalDL1_3229	250705737	0gi|298674424|ref|YP_003726174.1|	arCOG05282	2	V			Type I restriction-modification system methyltransferase subunit
HalDL1_Contig37	176868	G	A	A	0.975	F	F		159	0	4	0	163	ag									
HalDL1_Contig37	176902	G	A	A	0.99	F	F		191	0	2	0	193	gg									
HalDL1_Contig37	176942	T	C	C	0.969	F	F		0	31	0	1	32	tt									
HalDL1_Contig37	211164	A	G	G	0.947	T	T	HalDL1_3265	2	0	36	0	38	ta									
HalDL1_Contig37	219087	T	C	C	1	F	F		0	50	0	0	50	ct									
HalDL1_Contig37	250195	G	A	A	0.977	T	T	HalDL1_3308	85	0	2	0	87	gg	HalDL1_3308	250705744	9gi|313126061|ref|YP_004036331.1|	arCOG00805	1	L			ATP-dependent nuclease, subunit B
HalDL1_Contig37	280367	T	C	C	0.943	T	T	HalDL1_3327	0	50	0	3	53	gt	HalDL1_3327	250705746	8gi|292653982|ref|YP_003533880.1|	arCOG04041	1	K			Transcriptional regulator, contains wHTH domain
HalDL1_Contig37	280370	A	G	G	0.929	T	T	HalDL1_3327	4	0	52	0	56	ga	HalDL1_3327	250705746	8gi|292653982|ref|YP_003533880.1|	arCOG04041	1	K			Transcriptional regulator, contains wHTH domain
HalDL1_Contig37	280456	G	C	C	0.913	T	F	HalDL1_3328	0	105	2	8	115	cg	HalDL1_3328	250705746	9gi|292653576|ref|YP_003533472.1|	arCOG04518	1	K			Predicted transcriptional regulator, contains HTH domain
HalDL1_Contig37	280819	G	C	C	0.976	T	T	HalDL1_3328	1	160	3	0	164	gg	HalDL1_3328	250705746	9gi|292653576|ref|YP_003533472.1|	arCOG04518	1	K			Predicted transcriptional regulator, contains HTH domain
HalDL1_Contig37	282430	C	A	A	0.965	T	F	HalDL1_3329	139	5	0	0	144	gc	HalDL1_3329	250705747	0gi|292653584|ref|YP_003533480.1|	arCOG01510	1	L			Single-stranded DNA-binding replication protein A (RPA), large (70 kD) subunit or related ssDNA-binding protein
HalDL1_Contig37	283536	G	R	A	0.905	T	T	HalDL1_3330	19	0	2	0	21	tg									
HalDL1_Contig37	284700	T	Y	C	0.905	T	T	HalDL1_3331	0	19	0	2	21	tt									
HalDL1_Contig37	286619	C	Y	T	0.905	T	T	HalDL1_3333	0	2	0	19	21	tc									
HalDL1_Contig37	299291	A	G	G	0.959	T	F	HalDL1_3349	4	0	94	0	98	ga	HalDL1_3349	250705749	0gi|222476016|ref|YP_002564537.1|	arCOG04518	1	K			Predicted transcriptional regulator, contains HTH domain
HalDL1_Contig37	299293	C	T	T	0.93	T	T	HalDL1_3349	0	7	0	93	100	cc	HalDL1_3349	250705749	0gi|222476016|ref|YP_002564537.1|	arCOG04518	1	K			Predicted transcriptional regulator, contains HTH domain
HalDL1_Contig37	299596	A	M	C	0.917	T	F	HalDL1_3349	1	11	0	0	12	aa	HalDL1_3349	250705749	0gi|222476016|ref|YP_002564537.1|	arCOG04518	1	K			Predicted transcriptional regulator, contains HTH domain
HalDL1_Contig38	431234	C	T	T	1	T	F	HalDL1_0454	0	0	0	91	91	gc									
NC_012028	137835	T	C	C	0.906	T	F	Hlac_2875	0	48	0	5	53	gt	Hlac_2875	643706694	gi|222475799|ref|YP_002564320.1|	arCOG02751	1	L	COG3039	L	Transposase, IS5 family
NC_012028	137840	T	C	C	0.949	T	T	Hlac_2875	0	93	0	5	98	tt	Hlac_2875	643706694	gi|222475799|ref|YP_002564320.1|	arCOG02751	1	L	COG3039	L	Transposase, IS5 family
NC_012028	137911	C	T	T	0.992	T	F	Hlac_2875	0	6	1	858	865	tc	Hlac_2875	643706694	gi|222475799|ref|YP_002564320.1|	arCOG02751	1	L	COG3039	L	Transposase, IS5 family
NC_012028	137959	G	A	A	0.992	T	F	Hlac_2875	873	0	7	0	880	tg	Hlac_2875	643706694	gi|222475799|ref|YP_002564320.1|	arCOG02751	1	L	COG3039	L	Transposase, IS5 family
NC_012028	197570	A	G	G	0.921	F	F		8	0	93	0	101	ga									
NC_012028	200026	A	G	G	0.97	F	F		20	0	644	0	664	ca									
NC_012028	200046	C	T	T	0.95	F	F		0	33	0	621	654	ac									
NC_012028	200440	C	T	T	0.936	T	F	Hlac_2944	0	49	0	718	767	cc	Hlac_2944	643706759	gi|222475860|ref|YP_002564381.1|	arCOG04276	3	F	COG0209	F	Ribonucleotide reductase, alpha subunit
NC_012028	200757	A	G	G	0.944	T	T	Hlac_2944	41	0	691	0	732	ta	Hlac_2944	643706759	gi|222475860|ref|YP_002564381.1|	arCOG04276	3	F	COG0209	F	Ribonucleotide reductase, alpha subunit
NC_012028	200772	A	G	G	0.946	T	T	Hlac_2944	40	1	716	0	757	aa	Hlac_2944	643706759	gi|222475860|ref|YP_002564381.1|	arCOG04276	3	F	COG0209	F	Ribonucleotide reductase, alpha subunit
NC_012028	203299	T	C	C	0.93	T	F	Hlac_2946	0	621	0	47	668	tt	Hlac_2946	643706761	gi|222475862|ref|YP_002564383.1|	arCOG08098	1	K	COG1327	K	Transcriptional regulator NrdR (Zn-ribbon and ATP-cone domains)
NC_012028	208321	A	G	G	0.913	T	F	Hlac_2955	6	0	63	0	69	ga									
NC_012028	209696	C	T	T	0.914	T	T	Hlac_2958	0	8	0	85	93	gc	Hlac_2958	643706772	gi|222475873|ref|YP_002564394.1|	arCOG00467	1	L	COG1474	LO	Cdc6-related protein, AAA superfamily ATPase
NC_012028	209876	T	C	C	0.902	T	T	Hlac_2958	0	83	1	8	92	ct	Hlac_2958	643706772	gi|222475873|ref|YP_002564394.1|	arCOG00467	1	L	COG1474	LO	Cdc6-related protein, AAA superfamily ATPase
NC_012028	210644	T	C	C	0.917	F	F		0	77	0	7	84	at									
NC_012028	213155	G	A	A	0.937	T	F	Hlac_2962	89	0	6	0	95	ag	Hlac_2962	643706774	gi|222475875|ref|YP_002564396.1|	arCOG07767	4	S			Uncharacterized conserved protein
NC_012028	217727	A	G	G	0.946	F	F		5	0	87	0	92	ga									
NC_012028	221655	C	G	G	0.902	T	T	Hlac_2966	0	9	83	0	92	cc	Hlac_2966	643706778	gi|222475879|ref|YP_002564400.1|	arCOG04816	1	L	COG3505	U	TraG/TraD/VirD4 family enzyme, ATPase
NC_012028	221859	T	C	C	0.929	T	T	Hlac_2966	0	13	0	1	14	ct	Hlac_2966	643706778	gi|222475879|ref|YP_002564400.1|	arCOG04816	1	L	COG3505	U	TraG/TraD/VirD4 family enzyme, ATPase
NC_012028	221862	A	R	G	0.917	T	T	Hlac_2966	1	0	11	0	12	ca	Hlac_2966	643706778	gi|222475879|ref|YP_002564400.1|	arCOG04816	1	L	COG3505	U	TraG/TraD/VirD4 family enzyme, ATPase
NC_012028	222096	G	A	A	0.91	T	T	Hlac_2966	61	0	6	0	67	cg	Hlac_2966	643706778	gi|222475879|ref|YP_002564400.1|	arCOG04816	1	L	COG3505	U	TraG/TraD/VirD4 family

																									enzyme, ATPase
NC_012028	222816	G	A	A	0.98	T	T	Hlac_2966	50	0	1	0	51	cg	Hlac_2966	643706778	gi|222475879|ref|YP_002564400.1|	arCOG04816	1	L	COG3505	U	TraG/TraD/VirD4 family enzyme, ATPase
NC_012028	222873	T	C	C	0.947	T	T	Hlac_2966	0	71	0	4	75	ct	Hlac_2966	643706778	gi|222475879|ref|YP_002564400.1|	arCOG04816	1	L	COG3505	U	TraG/TraD/VirD4 family enzyme, ATPase
NC_012028	223262	C	T	T	0.931	T	F	Hlac_2966	0	6	0	81	87	gc	Hlac_2966	643706778	gi|222475879|ref|YP_002564400.1|	arCOG04816	1	L	COG3505	U	TraG/TraD/VirD4 family enzyme, ATPase
NC_012028	223374	C	T	T	0.935	T	T	Hlac_2966	0	6	0	87	93	gc	Hlac_2966	643706778	gi|222475879|ref|YP_002564400.1|	arCOG04816	1	L	COG3505	U	TraG/TraD/VirD4 family enzyme, ATPase
NC_012028	234391	T	A	A	0.95	T	F	Hlac_2976	19	0	0	1	20	at	Hlac_2976	643706786	gi|222475887|ref|YP_002564408.1|	arCOG06273	4	S			S-layer protein
NC_012028	238210	A	G	G	0.956	T	T	Hlac_2977	38	0	826	0	864	ta	Hlac_2977	643706787	gi|222475888|ref|YP_002564409.1|	arCOG00683	1	L	COG0675	L	Transposase
NC_012028	392982	C	A	A	0.956	T	F	Hlac_3155	194	8	1	0	203	gc									
NC_012028	475720	G	A	A	0.992	T	T	Hlac_3225	517	0	4	0	521	gg	Hlac_3225	643707021	gi|222476122|ref|YP_002564643.1|	arCOG04588	4	S			Uncharacterized conserved protein
NC_012029	33911	T	C	C	0.979	T	F	Hlac_0031	0	92	0	2	94	tt	Hlac_0031	643707090	gi|222478470|ref|YP_002564707.1|	arCOG02892	2	T	COG2766	T	Predicted Ser/Thr protein kinase
NC_012029	150233	T	C	C	0.916	T	F	Hlac_0141	0	87	0	8	95	ct	Hlac_0141	643707205	gi|222478580|ref|YP_002564817.1|	arCOG01706	3	C			Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide acyltransferase (E2) component or related enzyme
NC_012029	189693	A	G	G	0.963	F	F		4	0	105	0	109	ta									
NC_012029	229882	T	G	G	0.99	F	F		0	0	103	1	104	at									
NC_012029	232386	C	T	T	1	T	F	Hlac_0216	0	0	0	126	126	gc	Hlac_0216	643707279	gi|222478654|ref|YP_002564891.1|	arCOG04335	3	C	COG1301	C	Na+/H+-dicarboxylate symporter
NC_012029	259451	C	T	T	1	T	F	Hlac_0241	0	0	0	82	82	gc	Hlac_0241	643707304	gi|222478679|ref|YP_002564916.1|	arCOG01726	3	H	COG0142	H	Geranylgeranyl pyrophosphate synthase
NC_012029	302457	T	G	G	0.936	T	F	Hlac_0279	0	0	103	7	110	ct	Hlac_0279	643707342	gi|222478717|ref|YP_002564954.1|	arCOG02459	3	C	COG1527	C	Archaeal/vacuolar-type H+-ATPase subunit C
NC_012029	329952	T	G	G	0.988	F	F		0	0	84	1	85	gt									
NC_012029	334391	G	C	C	0.962	F	F		0	76	2	1	79	cg									
NC_012029	339315	T	G	G	0.976	T	T	Hlac_0317	0	0	83	2	85	ct	Hlac_0317	643707379	gi|222478754|ref|YP_002564991.1|	arCOG01338	3	C			Acyl-CoA synthetase, ATP-grasp containing subunit
NC_012029	345479	T	G	G	1	T	F	Hlac_0323	0	0	83	0	83	at	Hlac_0323	643707385	gi|222478760|ref|YP_002564997.1|	arCOG01296	2	O	COG0492	O	Thioredoxin reductase
NC_012029	384290	G	A	A	0.914	F	F		85	0	8	0	93	gg									
NC_012029	429985	G	S	C	0.909	T	F	Hlac_0412	0	10	1	0	11	gg	Hlac_0412	643707476	gi|222478848|ref|YP_002565085.1|	arCOG06273	4	S			S-layer protein
NC_012029	430668	A	G	G	0.944	T	F	Hlac_0412	1	0	17	0	18	ta	Hlac_0412	643707476	gi|222478848|ref|YP_002565085.1|	arCOG06273	4	S			S-layer protein
NC_012029	486012	A	C	C	0.912	T	F	Hlac_0468	6	62	0	0	68	ca	Hlac_0468	643707532	gi|222478903|ref|YP_002565140.1|	arCOG00498	4	R	COG0491	R	Metal-dependent hydrolase of the beta-lactamase superfamily II
NC_012029	494150	T	C	C	0.919	F	F		0	34	0	3	37	tt									
NC_012029	499574	C	T	T	0.963	T	F	Hlac_0480	0	1	0	26	27	gc	Hlac_0480	643707544	gi|222478915|ref|YP_002565152.1|	arCOG01740	2	U	COG0681	U	Signal peptidase I
NC_012029	561361	A	C	C	0.918	T	F	Hlac_0540	7	78	0	0	85	ga	Hlac_0540	643707605	gi|222478975|ref|YP_002565212.1|	arCOG00198	3	P	COG1120	PH	ABC-type cobalamin/Fe3+-siderophores transport system, ATPase component
NC_012029	614137	A	G	G	0.939	F	F		7	0	107	0	114	ta									
NC_012029	616396	A	C	C	0.984	T	F	Hlac_0596	2	124	0	0	126	ga	Hlac_0596	643707662	gi|222479031|ref|YP_002565268.1|	arCOG00517	4	R			Rhodanese Homology Domain fused to Zn-dependent hydrolase of glyoxylase family
NC_012029	649695	A	C	C	1	T	F	Hlac_0630	0	78	0	0	78	ca	Hlac_0630	643707696	gi|222479065|ref|YP_002565302.1|	arCOG00160	3	G	COG3833	G	ABC-type maltose transport system, permease component
NC_012029	863643	A	G	G	0.953	T	F	Hlac_0874	4	0	82	0	86	ta	Hlac_0874	643707939	gi|222479305|ref|YP_002565542.1|	arCOG04445	3	E			Acetolactate synthase, small subunit
NC_012029	864332	A	C	C	0.964	T	F	Hlac_0875	4	108	0	0	112	ta	Hlac_0875	643707940	gi|222479306|ref|YP_002565543.1|	arCOG01998	3	E	COG0028	EH	Acetolactate synthase large subunit or other thiamine pyrophosphate-requiring enzyme
NC_012029	865469	T	C	C	0.964	T	F	Hlac_0875	0	81	0	3	84	gt	Hlac_0875	643707940	gi|222479306|ref|YP_002565543.1|	arCOG01998	3	E	COG0028	EH	Acetolactate synthase large subunit or other thiamine pyrophosphate-requiring enzyme
NC_012029	878111	C	T	T	0.979	T	F	Hlac_0886	0	2	0	94	96	cc	Hlac_0886	643707951	gi|222479317|ref|YP_002565554.1|	arCOG00514	4	R	COG3467	R	Predicted flavin-nucleotide-binding protein
NC_012029	937444	T	C	C	0.974	T	F	Hlac_0938	0	38	0	1	39	gt	Hlac_0938	643708004	gi|222479369|ref|YP_002565606.1|	arCOG01324	3	F	COG2820	F	Uridine phosphorylase
NC_012029	958072	C	T	T	0.973	T	T	Hlac_0966	0	2	0	73	75	gc	Hlac_0966	643708032	gi|222479397|ref|YP_002565634.1|	arCOG00543	4	R	COG1782	R	Predicted metal-dependent RNase, consists of a metallo-beta-lactamase domain and an RNA-binding KH domain
NC_012029	1071355	T	C	C	0.963	T	T	Hlac_1071	0	181	0	7	188	at	Hlac_1071	643708138	gi|222479500|ref|YP_002565737.1|	arCOG10642	4	S			Uncharacterized conserved protein
NC_012029	1073958	G	C	C	0.949	T	T	Hlac_1074	0	223	12	0	235	cg	Hlac_1074	643708140	gi|222479502|ref|YP_002565739.1|	arCOG01376	2	M	COG0451	MG	Nucleoside-diphosphate-sugar epimerase
NC_012029	1123613	C	T	T	0.985	T	F	Hlac_1117	0	1	0	65	66	cc	Hlac_1117	643708182	gi|222479544|ref|YP_002565781.1|	arCOG00239	3	I	COG1024	I	Enoyl-CoA hydratase/carnithine

																								racemase
NC_012029	1174081	T	C	C	0.99	T	F	Hlac_1167	0	96	0	1	97	at	Hlac_1167	643708232	gi|222479594|ref|YP_002565831.1|	arCOG02280	1	K	COG3413	R	Predicted transcriptional regulator, C-terminal HTH-like domain
NC_012029	1175120	A	C	C	0.97	T	T	Hlac_1168	3	96	0	0	99	ta	Hlac_1168	643708233	gi|222479595|ref|YP_002565832.1|	arCOG01619	4	R	COG0656	R	Aldo/keto reductase, related to diketogulonate reductase
NC_012029	1218499	T	C	C	1	T	F	Hlac_1208	0	88	0	0	88	tt	Hlac_1208	643708273	gi|222479635|ref|YP_002565872.1|	arCOG02599	2	T	COG0784	T	Rec domain
NC_012029	1242166	A	G	G	0.958	T	F	Hlac_1231	4	0	92	0	96	aa	Hlac_1231	643708295	gi|222479656|ref|YP_002565893.1|	arCOG02369	2	T	COG0642	T	Signal transduction histidine kinase
NC_012029	1275454	T	C	C	0.968	T	T	Hlac_1263	0	90	0	3	93	ct	Hlac_1263	643708327	gi|222479688|ref|YP_002565925.1|	arCOG01430	3	E			Cysteine synthase
NC_012029	1329627	A	C	C	1	T	T	Hlac_1317	0	95	0	0	95	ta	Hlac_1317	643708381	gi|222479742|ref|YP_002565979.1|	arCOG02815	2	O	COG0229	O	Conserved domain frequently associated with peptide methionine sulfoxide reductase
NC_012029	1333200	A	G	G	0.96	T	F	Hlac_1320	4	0	97	0	101	aa	Hlac_1320	643708384	gi|222479745|ref|YP_002565982.1|	arCOG01580	1	K	COG1522	K	Transcriptional regulator, IclR family
NC_012029	1392755	G	C	C	0.986	T	F	Hlac_1382	0	72	1	0	73	tg	Hlac_1382	643708444	gi|222479804|ref|YP_002566041.1|	arCOG01018	1	J	COG0565	J	rRNA methylase
NC_012029	1418591	A	T	T	1	F	F		0	0	0	102	102	ca										
NC_012029	1439933	A	C	C	0.97	T	F	Hlac_1428	2	65	0	0	67	ta	Hlac_1428	643708489	gi|222479849|ref|YP_002566086.1|	arCOG04450	3	C	COG1668	CP	ABC-type Na+ efflux pump, permease component
NC_012029	1553325	C	T	T	0.949	F	F		0	4	0	75	79	gc										
NC_012029	1576719	A	G	G	0.971	T	T	Hlac_1556	3	0	99	0	102	ca	Hlac_1556	643708622	gi|222479976|ref|YP_002566213.1|	arCOG00874	1	K	COG1061	KL	ssDNA-dependent ATPase, helicase superfamily II
NC_012029	1606471	T	C	C	0.986	T	F	Hlac_1584	0	70	0	1	71	gt	Hlac_1584	643708650	gi|222480003|ref|YP_002566240.1|	arCOG07769	4	S			Uncharacterized conserved protein
NC_012029	1683580	T	C	C	0.91	T	F	Hlac_1662	0	91	0	9	100	at	Hlac_1662	643708730	gi|222480080|ref|YP_002566317.1|	arCOG01173	2	T			RecA-superfamily ATPase implicated in signal transduction
NC_012029	1708322	A	G	G	1	T	T	Hlac_1688	0	0	92	0	92	aa	Hlac_1688	643708757	gi|222480106|ref|YP_002566343.1|	arCOG01646	3	E	COG1506	E	Dipeptidyl aminopeptidase/acylaminoacyl-peptidase
NC_012029	1718840	T	C	C	0.969	F	F		0	93	0	3	96	tt										
NC_012029	1747552	T	C	C	0.965	T	F	Hlac_1731	0	82	0	3	85	at										
NC_012029	1777580	C	G	G	0.99	F	F		0	0	100	1	101	ac										
NC_012029	1804946	A	C	C	0.987	T	F	Hlac_1788	2	151	0	0	153	aa	Hlac_1788	643708854	gi|222480200|ref|YP_002566437.1|	arCOG03787	2	V			LEA14-like dessication related protein
NC_012029	1817550	C	T	T	0.972	F	F		0	3	0	105	108	gc										
NC_012029	1824288	T	A	A	0.969	F	F		94	0	0	3	97	tt										
NC_012029	1914793	A	G	G	1	F	F		0	0	107	0	107	ca										
NC_012029	1945308	A	G	G	0.979	T	F	Hlac_1946	2	0	95	0	97	ca	Hlac_1946	643709017	gi|222480357|ref|YP_002566594.1|	arCOG02365	2	T			Signal transduction histidine kinase with PAS domain
NC_012029	1971372	A	G	G	0.985	T	F	Hlac_1973	1	0	67	0	68	ca	Hlac_1973	643709044	gi|222480384|ref|YP_002566621.1|	arCOG02890	4	R	COG3608	R	Predicted deacylase
NC_012029	2004942	A	G	G	0.979	T	T	Hlac_2011	2	0	94	0	96	aa	Hlac_2011	643709082	gi|222480422|ref|YP_002566659.1|	arCOG01306	2	O	COG1222	O	ATP-dependent 26S proteasome regulatory subunit
NC_012029	2005576	T	A	A	0.921	T	F	Hlac_2011	93	0	0	8	101	gt	Hlac_2011	643709082	gi|222480422|ref|YP_002566659.1|	arCOG01306	2	O	COG1222	O	ATP-dependent 26S proteasome regulatory subunit
NC_012029	2031976	G	C	C	1	F	F		0	76	0	0	76	gg										
NC_012029	2032598	C	T	T	0.963	T	F	Hlac_2039	0	3	0	77	80	ac	Hlac_2039	643709110	gi|222480450|ref|YP_002566687.1|	arCOG01172	4	S	COG0467	T	Uncharacterized conserved protein
NC_012029	2045338	T	C	C	0.935	T	F	Hlac_2054	0	86	0	6	92	tt	Hlac_2054	643709125	gi|222480465|ref|YP_002566702.1|	arCOG03106	4	R	COG3631	R	Ketosteroid isomerase-related protein
NC_012029	2072517	A	C	C	0.962	T	F	Hlac_2084	3	76	0	0	79	ga	Hlac_2084	643709155	gi|222480495|ref|YP_002566732.1|	arCOG01576	3	P	COG2217	P	Cation transport ATPase
NC_012029	2122453	C	T	T	0.99	T	T	Hlac_2131	0	1	0	96	97	gc	Hlac_2131	643709201	gi|222480541|ref|YP_002566778.1|	arCOG01044	3	H	COG1648	H	Siroheme synthase (precorrin-2 oxidase/ferrochelatase domain)
NC_012029	2173707	G	C	C	0.925	F	F		0	37	3	0	40	gg										
NC_012029	2213035	T	C	C	1	T	F	Hlac_2227	0	46	0	0	46	gt	Hlac_2227	643709299	gi|222480637|ref|YP_002566874.1|	arCOG01166	1	L	COG0323	L	DNA mismatch repair enzyme (predicted ATPase)
NC_012029	2302147	A	G	G	0.959	T	F	Hlac_2311	3	1	116	1	121	ca	Hlac_2311	643709384	gi|222480720|ref|YP_002566957.1|	arCOG06225	3	C	COG2352	C	Phosphoenolpyruvate carboxylase
NC_012029	2414750	C	G	G	0.945	T	T	Hlac_2426	0	5	86	0	91	gc	Hlac_2426	643709499	gi|222480834|ref|YP_002567071.1|	arCOG04169	2	U	COG0201	U	Preprotein translocase subunit SecY
NC_012029	2494581	A	G	G	0.9	F	F		7	0	63	0	70	ta										
NC_012029	2576430	A	C	C	0.902	T	T	Hlac_2597	10	92	0	0	102	ga	Hlac_2597	643709670	gi|222481002|ref|YP_002567239.1|	arCOG00272	3	G	COG0697	GER	permease of the drug/metabolite transporter (DMT) superfamily
NC_012029	2579304	C	G	G	1	T	T	Hlac_2601	0	0	102	0	102	cc	Hlac_2601	643709674	gi|222481006|ref|YP_002567243.1|	arCOG00607	4	R	COG1994	R	Zn-dependent protease fused to CBS domain
NC_012029	2585143	C	T	T	0.951	T	F	Hlac_2606	0	4	0	77	81	cc	Hlac_2606	643709678	gi|222481010|ref|YP_002567247.1|	arCOG00625	3	P	COG2239	P	Mg/Co/Ni transporter MgtE (contains CBS domain)
NC_012029	2594549	C	T	T	0.987	T	F	Hlac_2613	0	1	0	77	78	gc	Hlac_2613	643709685	gi|222481017|ref|YP_002567254.1|	arCOG00914	3	E	COG4992	E	Ornithine/acetylornithine

																								aminotransferase
NC_012029	2600932	C	T	T	0.982	T	F	Hlac_2619	0	2	0	107	109	cc	Hlac_2619	643709691	gi|222481023|ref|YP_002567260.1|	arCOG00112	3	E	COG0137	E	Argininosuccinate synthase
NC_012029	2610263	A	G	G	0.919	F	F		8	0	91	0	99	ca										
NC_012029	2662169	G	C	C	0.972	T	T	Hlac_2675	0	104	3	0	107	cg	Hlac_2675	643709748	gi|222481079|ref|YP_002567316.1|	arCOG06335	4	S			Uncharacterized conserved protein
NC_012029	2708842	C	G	G	0.986	T	T	Hlac_2719	0	1	73	0	74	gc	Hlac_2719	643709794	gi|222481123|ref|YP_002567360.1|	arCOG01434	3	E	COG0498	E	Threonine synthase and cysteate synthase
NC_012030	55617	G	C	C	0.907	F	F		1	567	57	0	625	cg										
NC_012030	63886	G	T	T	1	F	F		0	0	0	22	22	gg										
NC_012030	67997	T	G	G	0.973	F	F		0	0	72	2	74	ct										
NC_012030	79998	C	T	T	1	T	F	Hlac_3329	0	0	0	385	385	tc	Hlac_3329	643709881	gi|222481211|ref|YP_002567447.1|	arCOG01445	2	V	COG1203	R	CRISPR-associated helicase Cas3
NC_012030	118390	T	C	C	0.958	T	T	Hlac_3363	1	346	0	14	361	gt	Hlac_3363	643709914	gi|222481244|ref|YP_002567480.1|	arCOG00683	1	L	COG0675	L	Transposase
NC_012030	320531	A	G	G	0.972	T	T	Hlac_3570	7	0	247	0	254	ca	Hlac_3570	643710108	gi|222481438|ref|YP_002567674.1|	arCOG04814	2	V			Type II restriction enzyme, methylase subunit
NC_012030	333711	C	T	T	1	F	F		0	0	0	6	6	gc										
NC_012030	341791	A	C	C	0.979	T	F	Hlac_3589	6	282	0	0	288	ca	Hlac_3589	643710127	gi|222481457|ref|YP_002567693.1|	arCOG02416	2	N			Predicted flagellin FlaG
NC_012030	422873	G	C	C	0.916	T	F	Hlac_3662	0	206	19	0	225	tg	Hlac_3662	643710193	gi|222481523|ref|YP_002567759.1|	arCOG06250	4	S			Uncharacterized conserved protein
NC_012030	423678	A	G	G	0.947	T	F	Hlac_3663	9	0	162	0	171	ta	Hlac_3663	643710194	gi|222481524|ref|YP_002567760.1|	arCOG06251	2	M			UDP-N-acetylmuramyl tripeptide synthase and Folylpolyglutamate synthase
NC_012030	424603	C	A	A	0.939	T	F	Hlac_3663	154	10	0	0	164	cc	Hlac_3663	643710194	gi|222481524|ref|YP_002567760.1|	arCOG06251	2	M			UDP-N-acetylmuramyl tripeptide synthase and Folylpolyglutamate synthase
NC_012030	425052	A	G	G	0.983	T	F	Hlac_3663	2	0	119	0	121	ta	Hlac_3663	643710194	gi|222481524|ref|YP_002567760.1|	arCOG06251	2	M			UDP-N-acetylmuramyl tripeptide synthase and Folylpolyglutamate synthase

Table S7. Fixed SNPs in tADL assigned to arCOGs.

Count	Functional category	Functional description	Class	Class description
17	R	General function prediction only	4	POORLY CHARACTERIZED
11	C	Energy production and conversion	3	METABOLISM
9	L	Replication; recombination and repair	1	INFORMATION STORAGE AND PROCESSING
8	J	Translation; ribosomal structure and biogenesis	1	INFORMATION STORAGE AND PROCESSING
6	E	Amino acid transport and metabolism	3	METABOLISM
6	G	Carbohydrate transport and metabolism	3	METABOLISM
6	P	Inorganic ion transport and metabolism	3	METABOLISM
5	S	Function unknown	4	POORLY CHARACTERIZED
5	T	Signal transduction mechanisms	2	CELLULAR PROCESSES AND SIGNALING
4	M	Cell wall/membrane/envelope biogenesis	2	CELLULAR PROCESSES AND SIGNALING
4	O	Posttranslational modification; protein turnover; chaperones	2	CELLULAR PROCESSES AND SIGNALING
3	F	Nucleotide transport and metabolism	3	METABOLISM
3	K	Transcription	1	INFORMATION STORAGE AND PROCESSING
2	N	Cell motility	2	CELLULAR PROCESSES AND SIGNALING
1	D	Cell cycle control; cell division; chromosome partitioning	2	CELLULAR PROCESSES AND SIGNALING
1	Q	Secondary metabolites biosynthesis; transport and catabolism	3	METABOLISM

Table S8. Similarity matrices for the primary replicons of tADL, DL31, *Hl* and DL1, and the "tADL-related 5th genome"[1].

TUD	5th_genome	DL1:Contig38	DL31:Contig115	Hl:NC_012029	tADL:Contig32
5th_genome	1				
DL1:Contig38	0.801	1			
DL31:Contig115	0.836	0.945	1		
Hl:NC_012029	0.862	0.834	0.777	1	
tADL:Contig32	0.959	0.779	0.858	0.785	1
ANIb	**5th_genome**	**DL1:Contig38**	**DL31:Contig115**	**Hl:NC_012029**	**tADL:Contig32**
5th_genome	1				
DL1:Contig38	0.697	1			
DL31:Contig115	0.699	0.715	1		
Hl:NC_012029	0.716	0.727	0.723	1	
tADL:Contig32	0.802	0.710	0.720	0.734	1

[1] Similarity matrices for BLASTN average nucleotide identity (ANIb) and tetranucleotide usage deviation (TUD) regression coefficients as determined by JSpecies for DL primary replicons. The 52 contigs attributed to "tADL-related 5th genome" are also included. For both metrics the tADL and "tADL-related 5th genome" are the most similar.

Table S9. Haloarchaeal genomes used for ANI and HIR analyses.

Short name	Long name
DL1	Halobacterium sp. DL1
DL31	halophilic archaeon DL31
tADL	halophilic archaeon True-ADL
H.lacusprofundi	Halorubrum lacusprofundi ATCC 49239
H.borinquense	Halogeometricum borinquense PR3, DSM 11551
H.hispanica	Haloarcula hispanica CGMCC 1.2049
H.jeotgali	Halalkalicoccus jeotgali B3, DSM 18796
H.marismortui	Haloarcula marismortui ATCC 43049
H.mediterranei	Haloferax mediterranei R-4, ATCC 33500
H.mukohataei	Halomicrobium mukohataei arg-2, DSM 12286
H.ruber	Halovivax ruber XH-70, DSM 18193
H.salinarum	Halobacterium salinarum R1, DSM 671
H.turkmenica	Haloterrigena turkmenica VKM B-1734, DSM 5511
H.utahensis	Halorhabdus utahensis AX-2, DSM 12940
H.volcanii	Haloferax volcanii DS2, ATCC 29605
H.walsbyi	Haloquadratum walsbyi HBSQ001, DSM 16790
H.walsbyi_C23	Haloquadratum walsbyi C23, DSM 16854
H.xanaduensis	Halopiger xanaduensis SH-6
Halo_NRC-1	Halobacterium sp. NRC-1
N.gregoryi	Natronobacterium gregoryi SP2, DSM 3393
N.magadii	Natrialba magadii ATCC 43099
N.occultus	Natronococcus occultus SP4, DSM 3396
N.pellirubrum	Natrinema pellirubrum 157, JCM 10476
N.pharaonis	Natronomonas pharaonis Gabara, DSM 2160
Nat_J7-2	Natrinema sp. J7-2

Table S10. Total length of HIR (bp) shared between the four DL genomes.

	tADL	**DL31**	***Hl***
DL31	6561		
Hl	47829	138307	
DL1	22569	68558	132560

Table S11. PCR primers used to amplify HIR.

Taxon + Shared region	Primer sequence
HalDL1_Contig37 20871..25836	ATAGACCTACACGAGAACACGACCAAG
HalDL1_Contig37 20871..25836	ACGCCCTACGAGACAGTGAGACAG
HalDL1_Contig37 20871..25836	GCCTGCTGCTTGGCGAGGAGTTC
HalDL1_Contig37 20871..25836	CGTGAGACGGTGCGAGGGTATG
HalDL1_Contig37 40156..44084	CCCCGATAGTAGTAATCAGAGGC
HalDL1_Contig37 40156..44084	GTAGAAATACGCAGTGGACGAACCC
HalDL1_Contig37 40156..44084	CAGTCTCCACAGGCGTTGATTC
HalDL1_Contig37 40156..44084	AGCAGTTTGTAGGCGGTAAGC
HalDL1_Contig37 44074..49706	CAGTCTCCACAGGCGTTGATTC
HalDL1_Contig37 44074..49706	AGCAGTTTGTAGGCGGTAAGC
HalDL1_Contig37 44074..49706	CGCCGTTGCCGTGAAGATG
HalDL1_Contig37 44074..49706	GTAGAGTTCGCCGAGCGTGATG
HalDL1_Contig37 49699..61930	CGCCGTTGCCGTGAAGATG
HalDL1_Contig37 49699..61930	GTAGAGTTCGCCGAGCGTGATG
HalDL1_Contig37 70308..74648	GGCTGACATCTAAGGCACTCGG
HalDL1_Contig37 70308..74648	GGCGGGACCTCAATCAACCAC
HalDL1_Contig37 70308..74648	GGCTGTATGGCGTGTTGTATTG
HalDL1_Contig37 70308..74648	CTTGTTCCAGAAGCGTCGTG
HalDL1_Contig37 101134..105279	GACGATGATACCAGCACCC
HalDL1_Contig37 101134..105279	GTTTCGCAACCAGATACGC
HalDL1_Contig37 102306..105542	CGTGTATCGGAATCATTGGAGGAG
HalDL1_Contig37 102306..105542	GGTTGGTTTCGTGAGCGTGGTC
HalDL1_Contig37 102306..105542	GATAACGGGTGACTCATACGCC
HalDL1_Contig37 102306..105542	GCAGGTCTCGCTGTCAGTGTTG
HalDL1_Contig38 453313..459149	GCAAGCCCCGACTAAGACAG
HalDL1_Contig38 453313..459149	CTGATGGTGAAGATGCTGACCG
HalDL1_Contig38 453313..459149	CTCCGATGAGACTCCCACTG
HalDL1_Contig38 453313..459149	CAGCGTGTTCAGGGCGTC
H.lac NC_012028 7615..29467	CCAGAATCAGAGACATCGCTCAAG
H.lac NC_012028 7615..29467	GTAGTAGTATCTGTAGTACCTCGGCAC
H.lac NC_012028 7615..29467	GCGGGAAGATCAGTGAGTACGAC
H.lac NC_012028 7615..29467	GAAGAGTAGTGGGAACGACGGC
H.lac NC_012028 54359..60782	CGGCGTTCTCAGGTTCTTCG
H.lac NC_012028 54359..60782	GCTAAGATAGTACAGTGCCGTGG
H.lac NC_012028 54359..60782	GCCCGCAATGACGAAGAC
H.lac NC_012028 54359..60782	GACGGCTTCTTCAGATCCCC
DL31_Contig114 59595..76202	GGCTGGGCTGGAACGAGAC
DL31_Contig114 59595..76202	GGTAGTGCTACGCTAAAACAGTGCC
DL31_Contig114 59595..76202	GGACTACGGTGGCAATCTCTATCTAATG
DL31_Contig114 59595..76202	GAGGAGGCTATAAATGGTAGAATCGGG
DL31_Contig114 111139..146012	GCGTCGGAAATCAGTGCTGAG
DL31_Contig114 111139..146012	CCTCAGAGAGTTACACGTCCATCC
DL31_Contig114 664513..683577	CCCAACCCCACCGTTTTG
DL31_Contig114 664513..683577	GTATGATGGATCGTTGACCTCGG
DL31_Contig114 664513..683577	CGATACTAATGTCCTCACTCAACTGG
DL31_Contig114 664513..683577	CACTCTCACCGTCTCGTCC
True-ADL 1940216..1952775	GCCGATGTTCCGAGAGGTTCAG
True-ADL 1940216..1952775	GTCGTCGTTCTGCTGGAGGC
True-ADL 1940216..1952775	CGGCGTGTAATGGCAACTGGCAC
True-ADL 1940216..1952775	CACGGCTGTCAGAGTGTCC
True-ADL 1234151..1243860	GTCCAACTCGCAACACTCGG
True-ADL 1234151..1243860	GTACTCAGCCACATAGGGTCC
True-ADL 1234151..1243860	GGACCTCACCCATACCACG
True-ADL 1234151..1243860	GTGCTCACTCCGATAATTCCTGC
True-ADL 1243876..1250380	GCCCAAGTGTAGCCGTCATC
True-ADL 1243876..1250380	CTGACTTGAGTACGACGCTGG

References

1. Denef, V. J. & Banfield, J. F. (2012) In situ evolutionary rate measurements show ecological success of recently emerged bacterial hybrids. *Science* 336, 462–466.
2. Franzmann, P. D. *et al.* (1988) *Halobacterium lacusprofundi* sp. nov., a halophilic bacterium isolated from Deep Lake, Antarctica. *Syst. Appl. Microbiol.* 11, 20–27.
3. Ferris, J. M. & Burton, H. R. (1988) The annual cycle of heat content and mechanical stability of hypersaline Deep Lake, Vestfold Hills, Antarctica. *Hydrobiologia* 165, 115–128.
4. Mandach, von, C. & Merkl, R. (2010) Genes optimized by evolution for accurate and fast translation encode in Archaea and Bacteria a broad and characteristic spectrum of protein functions. *BMC Genomics* 11, 617.
5. Capes, M. D., DasSarma, P. & DasSarma, S. (2012) The core and unique proteins of haloarchaea. *BMC Genomics* 13, 39.
6. Touchon, M. & Rocha, E. P. (2007) Causes of insertion sequences abundance in prokaryotic genomes. *Mol Biol Evol* 24:969–981.
7. Pina, M., Bize, A., Forterre, P. & Prangishvili, D. (2011) The archeoviruses. *FEMS Microbiol. Rev.* 35, 1035–1054.
8. Pietila M. K. *et al.* (2012) Virion architecture unifies globally distributed pleolipoviruses infecting halophilic archaea. *J. Virol.* 86, 5067-5079.
9. Walsby, A. E. (1994) Gas vesicles. *Microbiol Rev* 58:94-144.
10. Bardavid, R. E., Khristo, P., & Oren, A. (2008) Interrelationships between *Dunaliella* and halophilic prokaryotes in saltern crystallizer ponds. *Extremophiles* **12**, 5-14.
11. Gibson, J. A. E. (1999) The meromictic lakes and stratified marine basins of the Vestfold Hills, East Antarctica. *Antarct Sci* 11:175–192.
12. Ng, C. *et al.* (2010) Metaproteogenomic analysis of a dominant green sulfur bacterium from Ace Lake, Antarctica. *ISME J* 4:1002–1019.
13. Lauro, F. M. *et al.* (2011) An integrative study of a meromictic lake ecosystem in Antarctica. *ISME J* 5:879–895.
14. Benlloch, S. *et al.* (2001) Archaeal biodiversity in crystallizer ponds from a solar saltern: culture versus PCR. *Microb Ecol* 41:12-19.
15. Ochsenreiter, T., Pfeifer, F. & Schleper, C. (2002) Diversity of Archaea in hypersaline environments characterized by molecular-phylogenetic and cultivation studies. *Extremophiles* 6: 267-274.
16. Oh, D., Porter, K., Russ, B., Burns, D. & Dyall-Smith, M. (2010) Diversity of *Haloquadratum* and other haloarchaea in three, geographically distant, Australian saltern crystallizer ponds. *Extremophiles* 14:161-169.
17. Rusch, D. B. (2007) *et al.* The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. *PLoS Biol.* 5, e77.
18. Yau, S. *et al.* (2011) Virophage control of antarctic algal host-virus dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6163–6168.
19. Brown, M. V. *et al.* (2012) Global biogeography of SAR11 marine bacteria. *Mol. Syst. Biol.* 8, 595.
20. Wilkins, D. *et al.* (2012) Biogeographic partitioning of Southern Ocean microorganisms revealed by metagenomics. *Environ. Microbiol.* 15, 1318–1333.

21. Williams, T. J. *et al.* (2012) The role of planktonic Flavobacteria in processing algal organic matter in coastal East Antarctica revealed using metagenomics and metaproteomics. *Environ. Microbiol.* 15, 1302–1317.
22. Yau, S., *et al*. (2013) Metagenomic insights into strategies of carbon conservation and unusual sulfur biogeochemistry in a hypersaline Antarctic lake. *ISME J.* doi:10.1038/ismej.2013.69
23. Burns, D. G., Camakaris, H. M., Janssen, P. H. & Dyall-Smith, M. L. (2004) Cultivation of Walsby's square haloarchaeon. *FEMS Microbiol. Lett.* 238, 469–473.
24. Burns, D. G. *et al.* (2007) Haloquadratum walsbyi gen. nov., sp. nov., the square haloarchaeon of Walsby, isolated from saltern crystallizers in Australia and Spain. *Int. J. Syst. Evol. Microbiol.* 57, 387–392.
25. Dyall-Smith M.L. (2009) The Halohandbook: Protocols for halobacterial genetics. In., 7.2 edn: http://www.haloarchaea.com/resources/halohandbook/.
26. Margulies M. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437: 326-327
27. Bennett S. (2004) Solexa Ltd. Pharmacogenomics. 5:433-8.
28. Engelbrektson A, Kunin V, Wrighton KC, Zvenigorodsky N, Chen F, Ochman H *et al* (2010). Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *ISME J* 4: 642–647.
29. Kunin V, Engelbrektson A, Ochman H, Hugenholtz P (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ Microbiol* 12: 118.
30. Zerbino, D. and E. Birney. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Research 18: 821-829.
31. Ludwig, W. *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Res.* 32, 1363–1371.
32. McDonald, D. *et al.* (2012) An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* 6, 610–618.
33. Niu, B., Zhu, Z., Fu, L., Wu, S. & Li, W. (2011) FR-HIT, a very fast program to recruit metagenomic reads to homologous reference genomes. *Bioinformatics* 27, 1704–1705.
34. Bastian, M., Heymann, S. & Jacomy, M. (2009) Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the Third International ICWSM Conference*.
35. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.
36. Johnson, P. L. & Slatkin, M. (2009) Inference of microbial recombination rates from metagenomic data. *PLoS genetics* 5, e1000674.
37. Wysoker, A., Tibbetts, K., Fennell, T. Picard. (2009) http://picard.sourceforge.net/.
38. Gordon, D., Abajian, C. & Green, P. (1998) Consed: a graphical tool for sequence finishing. *Genome Res.* 8, 195–202.
39. Canty, A., Ripley, B., (2013) Bootstrap functions. http://cran.r-project.org/web/packages/boot/.

40. Sharp, P. M. & Li, W. H. (1987) The codon Adaptation Index – a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15, 1281–1295.
41. Peden, J. F. (2005) Correspondence Analysis of Codon Usage. http://codonw.sourceforge.net/.
42. Li, L., Stoeckert, C. J. & Roos, D. S. (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189.
43. Makarova, K. S., Sorokin, A. V., Novichkov, P. S., Wolf, Y. I. & Koonin, E. V. (2007) Clusters of orthologous genes for 41 archaeal genomes and implications for evolutionary genomics of archaea. *Biol. Direct* 2, 33.
44. Zhang, Z. *et al.* (2006) KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics* 4, 259–263.
45. Allen, M. A. *et al.* (2009) The genome sequence of the psychrophilic archaeon, *Methanococcoides burtonii*: the role of genome evolution in cold adaptation. *ISME J* 3:1012–1035.
46. Fraley, C. & Raftery, A. E. (2006) Mclust: an R package for normal mixture modeling. http://www.stat.washington.edu/mclust/.
47. Galardini, M., Biondi, E. G., Bazzicalupo, M. & Mengoni, A. (2011) CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. *Source Code Biol. Med.* 6, 11.
48. DeMaere, M. Z., Lauro, F. M., Thomas, T., Yau, S. & Cavicchioli, R. (2011) Simple high-throughput annotation pipeline (SHAP). *Bioinformatics* 27, 2431–2432.
49. Richter, M. & Rosselló-Móra, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131.
50. Edgar, R. C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 113.
51. Tillett, D. & Neilan, B. A. (2000) Xanthogenate nucleic acid isolation from cultured and environmental cyanobacteria. *J. Phycol.* 36, 251–258.
52. Wilkins, D. *et al.* Key microbial drivers in Antarctic aquatic environments. *FEMS Microbiol. Rev.* (2012). doi: 10.1111/1574-6976.12007.
53. Zwartz, D., Bird, M., Stone, J. & Lambeck, K. Holocene sea-level change and ice-sheet history in the Vestfold Hills, East Antarctica. *Earth Planet. Sci. Lett.* **155,** 131–145 (1998).
54. Bowman, J. P. J., McCammon, S. A. S., Rea, S. M. S. & McMeekin, T. A. T. The microbial composition of three limnologically disparate hypersaline Antarctic lakes. *FEMS Microbiol. Lett.* **183,** 81–88 (2000).
55. Ferris, J. M. & Burton, H. R. The annual cycle of heat content and mechanical stability of hypersaline Deep Lake, Vestfold Hills, Antarctica. *Hydrobiologia* **165,** 115–128 (1988).
56. Krzywinski, M. *et al.* (2009) Circos: an information aesthetic for comparative genomics. *Genome Res* 19:1639–1645.